South African Journal *of* Science

volume 117
number 9/10

Where are rhinos safest?

UN's Decade of Ocean Science for Sustainable Development

Home gardening to reduce food insecurity during economic distress

Excessive red tape is strangling biodiversity collections and research in South Africa

Driving mechanisms of cold water temperature anomalies on the Sodwana reefs

POPIA compliance in research involving children

ASSAf
ACADEMY OF SCIENCE OF SOUTH AFRICA

Volume 117
Number 9/10
September/October 2021

South African Journal of Science

eISSN: 1996-7489

Published by
the Academy of Science of South Africa (www.assaf.org.za) with financial assistance from the Department of Science & Innovation.

Design and layout
SUN MeDIA Bloemfontein t/a SunBonani Media
T: 051 444 2552
E: publish1@sunbonani.co.za

Correspondence and enquiries
sajs@assaf.org.za

Submissions
Submissions should be made at www.sajs.co.za

Review Articles

Research Articles

Cover caption
Colourful corals. Corals become stressed when exposed to elevated temperatures for extended periods and bleaching can occur. In an article on page 70, Wells and colleagues analyse temperature measurements taken between 1994 and 2015 on Nine Mile Reef at Sodwana to elucidate the driving mechanisms of the temperature anomalies around Sodwana. Sodwana experiences short-term temperature fluctuations that may provide relief from coral bleaching and be crucial in the future survival of the Sodwana reef system.

Insiders and outsiders

Every week, the *South African Journal of Science* receives submissions which are clearly not within the scope of the Journal. As our mission statement says, as a multidisciplinary journal, we are interested in 'publishing high-quality original research from Africa or on African-relevant issues'. From an administrative point of view, dealing with these out-of-scope submissions is quite easy, if time-consuming. We can simply inform authors of the scope issue and suggest that they submit elsewhere.

Underlying this common occurrence, though, is information about the current global climate in academic publishing, and this is an issue of broader concern. Almost none of these submissions are from scholars in high-income countries. Another reasonably common occurrence for us is a pre-submission query in which there is inevitably a question about page fees payable to the Journal, also invariably from a scholar who is not in a high-income country. Our Journal is in the very lucky position of being a diamond/platinum open access journal, meaning that we do not have article-processing charges because we receive funding from the Department of Science and Innovation through ASSAf, so the query is easy to deal with, but is revealing in itself.

What is clear from our experiences, and these are far from unique to our Journal, is that there is a cohort of scientists and scholars based in low- and middle-income countries (and certainly not in Africa alone) who are trying to have their work published in journals which enjoy a good reputation (as measured, for example, by accredited impact factors), and read by a global audience. A further feature of these submissions, and not just of these submissions, is that many authors struggle to write in clear English. Where the quality of written English is good, though, there are many occasions where the authors do not appear to have thought adequately about the multidisciplinary audience of the Journal and write in ways which may exclude our potential readership from being able easily to engage with the research being presented.

We are certainly not alone in the science community in South Africa and further afield in grappling with the imperative to open science in any way that we can in terms of who produces knowledge and who has access to it. At the same time, though, every week we gain a glimpse into the worlds of those who continue to be excluded. Although we at the Journal do our best to assist all those who wish to publish with us and to suggest alternative outlets, the fact is that, generally speaking, submissions which come from institutions and groups with established strong research writing traditions get an easier path to actually getting to the point of being peer reviewed. Anonymous and fair peer review is central to our work, but it seems to be the case that authors who are well networked in the science community may have directly or indirectly received the support necessary to be able to present to a journal editor a document for review which is in the correct format, within the journal scope, and ready to be assigned for peer review. It is our impression that it is generally these authors as well who will be familiar with conventions around plagiarism, and who will be less likely to have manuscripts returned for changes on the basis of similarities to already published work. The odds, then, seem to be stacked in favour of those with a degree of what may be termed academic social capital – linkages with those who understand the system and, indeed, may have the power to modify and enforce its rules.

The paradox in all of this is, of course, clear. On the one hand, our Journal is committed to practices of inclusion and openness, recognising that for science to be strong we need wide and diverse participation and to welcome a diversity of voices and approaches. On the other hand, despite the clear progress that the science community is making to diversify, we may be experienced by some who wish to publish with us as belying our commitment to inclusivity by not allowing them entrance, and not fully supporting them in becoming part of our or related networks. It is clearly beyond the resources of a Journal like ours to solve the global problems of unequal access to educational opportunities and research resources, the dominance of English as an international language of science, and, indeed, broader contributory questions linked to global poverty and inequality. But we do get a sense of the efforts excluded people are going to in order to try to be included.

Part of the problem, here, lies of course in the often-discussed commodification of research and research outputs, and the impact of metrics and audit culture on academic life globally. We believe that it is important for all scientists to engage with critiques of an output-driven system and with the debates concerning the limitations of metrics like impact factors and h-indices, for example, to create a universally applicable assessment of quality. But as a Journal, we see, possibly more than others may, some of the consequences of these factors for academic writing. It appears from some of those trying to gain access to the journal (and possibly for some of those who succeed to publish with us), that publications may be viewed as products in themselves, items to be counted and ticked off, used as materials to give access to jobs, grants, promotions and other opportunities. It is indeed the case that publications act as commodities in this way. But what may be being lost here is the role of writing as part of the research process itself.

One of the challenges and opportunities of writing for a multidisciplinary journal rests precisely in the fact that authors have to think about the audience, and to work at making their own knowledge accessible to those who may not share their assumptions or methodological training and expertise. The adage that the best way to learn is to teach has relevance here. We hope our authors gain from publishing with us more than just through racking up citations or burnishing curricula vitae (not that these are trivial issues in the lives of researchers and academics). We hope that in the process of doing the difficult work of writing for our diverse audience, authors also benefit from the rigour involved in trying to make complex ideas as broadly accessible as possible. This takes clarity and skill, and our Journal is very lucky to be able to publish work which demonstrates this clarity and skill. We have an extra ask of ourselves and of our readership, though. All those who are lucky enough, for a range of reasons, to be on the inside of the science community as it currently exists have a responsibility and an opportunity to broaden access to that community. Some of this, we suggest, is through the support of developing clear, accessible academic writing.

HOW TO CITE:
Swartz L. Insiders and outsiders. S Afr J Sci. 2021;117(9/10), Art. #12345. https://doi.org/10.17159/sajs.2021/12345

AUTHOR:
Ditshego Masete[1]

AFFILIATION:
[1]Faculty of Humanities, Social Sciences and Education, University of Venda, Thohoyandou, South Africa

CORRESPONDENCE TO:
Ditshego Masete

EMAIL:
Ditshego.masete@univen.ac.za

HOW TO CITE:
Masete D. How do we know if and when science makes a difference? S Afr J Sci. 2021;117(9/10), Art. #12226. https://doi.org/10.17159/sajs.2021/12226

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
National Science Week, science engagement, science for society, impact of science

PUBLISHED:
29 September 2021

How do we know if and when science makes a difference?

The Academy of Science of South Africa and the *South African Journal of Science* hosted a webinar on 4 August 2021 as part of their participation in South Africa's National Science Week – the annual celebration of science, engineering and technology led by the Department of Science and Innovation. The webinar was entitled: How do we know if and when science makes a difference? The panellists discussed the impact and role of scientists and science in contemporary society.

The panellists included Mwazvita Dalu, an NRF Innovation Postdoctoral Research Fellow at Unisa and Research Associate with the Open University, UK; Ernest Dube, an agronomy research scientist and Senior Lecturer in the School of Natural Resource Management at Nelson Mandela University; and Jennifer Fitchett, an Associate Professor of Physical Geography at the University of the Witwatersrand.

The programme director, Leslie Swartz, a Professor of Psychology at Stellenbosch University and Editor-in-Chief of the *South African Journal of Science*, stressed the importance of the relationship between science and society in his opening. He drew the attention of the audience and the presenters to the interest sparked by this webinar in calling upon scientists to have a conversation on the impact of science on society.

Swartz emphasised that, given the narratives circulating in the public sphere about the COVID-19 pandemic and vaccines, it is indeed important that science be questioned and that scientists be called upon to enlighten society on the difference that science makes in our lives to restore society's trust in science. This is a difficult period for science, and its credibility, because there was once a period in history when science held all authority and was not questioned, Swartz told the audience.

Interestingly, Dalu opened her presentation with reference to the feature film *The Day After Tomorrow* (2004). The film is about an apocalypse which results in a new world covered with ice. This is apt at a time when we are not sure how the world and life post-COVID pandemic will be – will we go back to life as we know it, and if not, what will the 'new normal' look like? Dalu related the narrative of the film to the context of the webinar as support of scientific findings that could make the world a better place for all. In this fictional film, the government, which was initially reluctant to support science, comes to the realisation that science is active and could be used to humanity's advantage. Although fictional, the film conveys the critical question of science and society by showing how humanity can benefit from scientific discoveries. It alludes to science being active and making a difference.

The purpose of science is to better society, Dalu told the participants. As such there is a need to establish clear, impactful and accessible communication channels through which scientists can see the difference they make in society. Essentially, science is for sustainability in the future. The messages packaged 'by science' for society must reach their audience and be deconstructed and easily understood in order to enhance the achievement of a scientific difference.

Currently, scientists see their contributions to society through the number of citations and downloads of their publications. These two metrics, among others, indicate that society is interested in the messages communicated by scientists for various reasons. Citations to scientific discoveries and recommendations in policy documents also indicate that, in the long term, ideas which emerge through scientific engagement, recommendations and discoveries help to shape the world and the future.

Seemingly, the advancement of social media and the mass media has allowed science to come to the doorstep of the public. These tools have made science more accessible. Dalu acknowledged social media platforms as some of the key communication channels through which scientists and society engage to make science a social practice. She acknowledged that the impact factor does not account for these interactions and suggested that the recent relinquishing of the journal impact factor as a measure of performance by some universities will open the space for 'good science' without compromising its quality by populist tendencies associated with career advancement.

On a different note, Dube spoke to the audience about agronomy and food security in South Africa. He maintained that current food production in South Africa is mostly driven by agronomy research. Agronomy is, according to Dube, 'the application of different scientific methodologies for the improvement and management of major stable food crops'. These include wheat and maize in South Africa. He also brought to the attention of the audience the flexibility of agronomy in integrating many other science disciplines such as soil science, ecology, entomology, climatology, sociology and agricultural economics to achieve its greater goals.

The success of farmers is directly linked to their implementation of scientific discoveries and recommendations from agronomy researchers. The advantage of agronomy research is that it is practical and therefore its effects are felt immediately. This suggests that agronomy as a field of science actively participates in making the world a better place. Dube demonstrated that agronomy researchers deal directly with farmers, therefore their effects are seen immediately, unlike in other disciplines where the impact may be felt only later. He emphasised that agronomy research is active and empirical.

Dube further demonstrated the effects of agronomy research in our everyday life. He reminded the audience that the food they had eaten that day was produced with the help of agronomy research. He also noted that – due to various factors such as the constant population growth, climate change, shortage of fertile land for farming,

drought, and the excessive use of pesticides – there is an urgent need to discover and implement various scientific methods to directly answer the call for food security.

His urgency is accurate given that statistics show that almost double the current global food supply will be needed in 2050 to feed a population of approximately 9 billion people.[1] However, the alarming rate at which the global population is increasing may mean that there could be a shortage of land, water and energy required to feed people and agronomy research should be employed to address this challenge.

To Dube, this challenge is fuelled by the current spike in urbanisation and the increased consumption of wheat. He argued that urbanisation has increased the change in people's diets. Currently, South Africa produces 2 million tons of wheat but consumes 3.2 million tons per annum. This suggests that there is an overall shortfall of 1.2 million tons. Therefore, conservation agriculture should be adopted to cover this shortfall while also conserving the environment. This method of farming must also be adopted to help reduce farmers' input costs, and to keep food safer through reducing the use of pesticides.

Dube concluded his presentation by illustrating how the agricultural value chain contributes significantly and directly to life and the economy through employment of farm and mill workers. This must indeed be seen as a qualitative and tangible difference made by science.

Fitchett, the last presenter, opened her presentation by posing the question 'When does science make/not make a difference'? She then went on to draw the attention of the audience and the other presenters to yet another crucial question: 'What does difference mean?' In essence she asked, what is this 'scientific difference' that is being talked about? To further interrogate the question, she posited that 'science always makes a difference' and centred her discussion around this sentiment.

In substantiating her thesis Fitchett argued that:

> *The pursuit and production of new knowledge and the process involved therein and the engagement with knowledge as a tangible and often growing outcome is something in itself that is inherently making a difference.*

She maintained that any work done in science, whether in the social sciences or any other field, is a stepping stone to future breakthroughs. If such a premise is accepted, then even if we think that one particular piece of work is not groundbreaking, the fact that it was done as research suggests that there was a need, and as such, the work will ultimately make a difference. This is because it will at some point allow someone else to build on it to enable breakthroughs and developments.

In some way echoing the words of Dalu on the impact factor, Fitchett maintained that science makes a difference when we see breakthroughs and discoveries, solve societal ills and raise awareness of problems. She further argued that society's understanding of issues and the world is often driven by science, and these should be seen as quantifiable results of science. Like Dalu, Fitchett calls for quantitative markers of science's contribution to society other than the impact factor for fair judgement of impact.

For Fitchett, the very process of conducting science and of being engaged in scientific research always makes a difference. She told the audience that society is able to see the inherent difference that science makes, through the process and the output which includes, but is not limited to, the training of researchers, the communication of scientific research findings through various means and platforms such as publications, webinars, conferences, public lectures and other communicable methods.

Interestingly, Fitchett acknowledged that science is a coin with two sides: just as it can do good, it can also be detrimental and catastrophic. Some of the major challenges in the world today, such as global warming, are partially an aftermath of scientific advancement. She asserted that today's solutions might be detrimental in the future, and that what may seem insignificant now, could be impacting on other systems and may have a greater impact years down the line – the butterfly effect. To come back to her initial question on what scientific difference is, Fitchett suggested that the difference is not always positive. But whether it is good or bad, science always makes a difference.

Fitchett further acknowledged that scientific difference is subjective. It is dependent on research institutions, universities and personal conventions. It is dependent on what researchers see and hold as important and worth pursuing. Therefore, scientific difference should not be viewed narrowly or simplistically.

In concluding her presentation, she commented on society's participation in science. Fitchett suggested that society should ask itself how it can support science and should never regard any output as insignificant and not impactful. There is a need to encourage and support 'science for good' organisations and help them in achieving their goals. It is important for society to help scientists to see the difference they are making and to encourage them to celebrate these differences. In a way, Fitchett called for a circular communication model of science in which scientists get feedback on how their science impacts society.

Clearly the presenters all share a similar sentiment about the vigor of science. They all maintained that science is not just a philosophical theory, but an empirical subject and an active social practice.

The National Science Week webinar entitled 'How do we know if and when science makes a difference?' can be viewed at https://youtu.be/4JBRGTE12UY

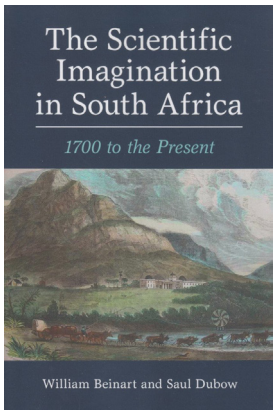
Competing interests

I have no competing interests to declare.

Reference

1. Whitfield B. The upside of down: How chaos and uncertainty breed opportunity in South Africa. Johannesburg: Pan Macmillan; 2021.

BOOK TITLE:
The scientific imagination in South Africa. 1700 to the present



AUTHORS:
William Beinart
Saul Dubow

ISBN:
9781108945578 (ebook, 418 pp, USD27.60)
9781108940085 (softcover, 418 pp, USD33.98)
9781108837088 (hardcover, 418 pp, USD83.49)

PUBLISHER:
Cambridge University Press, Cambridge, UK

PUBLISHED:
2021

REVIEWER:
Wieland Gevers[1]

AFFILIATION:
[1]Emeritus Professor of Medical Biochemistry, University of Cape Town, Cape Town, South Africa

EMAIL:
wieland.gevers@uct.ac.za

HOW TO CITE:
Gevers W. A well-told history of science in South Africa. S Afr J Sci. 2021;117(9/10), Art. #11967. https://doi.org/10.17159/sajs.2021/11967

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

PUBLISHED:
29 September 2021

A well-told history of science in South Africa

It is wonderful when the mere title of a book already contains the ingredients for an interesting debate in one's mind. Is the specifically scientific imagination different from other forms of imagination? Can 'science' be done without imagination? Realising this, the authors of this extremely readable and informative book have discussed their choice of title and focus on page 3, starting with a definition of 'scientific imagination' as 'an expression of human curiosity, ingenuity, and the ability to make unlikely connections'. They go on to consider the role of the imagination in the inductive scientific method as commonly understood and practised, and the question as to how inventions and innovation come about. They admit to being particularly interested in the 'imbrication' of science with the shaping of society in economic and political terms. (Later in the book, one encounters something they call an 'imaginary'– a widely shared rather than individual imagining of a possible future brought about by scientific and technological progress.)

William Beinart and Saul Dubow are South African expatriates working at Oxford and Cambridge Universities, respectively; they have both made splendid careers in the UK while retaining firm intellectual-academic connections with South Africa. Their new book is a pleasure to read; it is well paced and organised, and highly informative despite covering a good deal of ground. The chapters are imaginatively named and framed in their main contexts. It is a *tour de force* of accessible and meaningful history. For South Africans who believe that the past is a reasonable guide to the future, I would say it is essential reading. (This is not to say that there are not any gaps – see below – or that the index is not a little annoyingly incomplete.)

The book begins with an introductory chapter that explores some of the considerations that have guided the authors in their choice of material and emphasis. They begin inter alia with the core problem in a colonised country of how indigenous or vernacular (orally) transmitted knowledge can approximate to systematic (published) science. They argue that because 'all knowledge organises or systematises information or observations', there are 'no hard and fast boundaries between science and indigenous or local knowledge'. Yet they write largely about 'published or written ideas by those with specialist training who were familiar with disciplinary developments and international thought'. From the point of view of practical history-writing about the 'dead' past, that is understandable, if one relies only on evidence available about vernacular or local knowledge in accounts left by scientifically informed observers (one omission in the book is the extensive work done by South African archaeologists on the origin and diffusion of indigenous metallurgy and farming). Beinart and Dubow make a point of including such descriptions, although few of the writers concerned were curious as to how transmission/dissemination and replication of knowledgeable practice took place, especially that concerning plant materials, or how much of the placebo effect underlay the apparent success of observed treatments and other techniques. It is gratifying that trained scientists such as Peter Kolb and Anders Sparrman were notably more sympathetic towards the knowledge spheres of the indigenes they encountered than were persons untrained in objective observation and analysis. It becomes clear from the book that the human mind adapts to situation and context in a remarkable variety of ways, and that individual and group survival requires the exercise of the brain in what may justifiably be called 'adaptive savantism', something which may even apply to the offshoots of mainstream Western science in a country at the margins like South Africa.

The authors are careful not to include excessive biographic detail but expand when key figures are described in their specific contexts. The 18th century field science 'scouts' who travelled extensively in Dutch-ruled South Africa and reported its wonders to an ever-more curious Europe are deservedly given good coverage, as are both the eminent visitors (John Herschel and Charles Darwin) and more settled notables of the British-ruled 19th century. The rise of colonial scientific institutions led to the kinds of 'imaginaries' in which shared aspirations were added to the individual efforts of a relatively small number of multi-talented locals such as Andrew Bain and William Atherstone who began to sow the seeds of later areas of major focus. The 'wake-up call' of the discoveries of massive sources of diamonds and gold from 1870 onwards were accompanied by science-based technological innovation in geology (prospecting) and mining, as well as agriculture. Sheep and ostrich farming were surprisingly significant in the latter regard, apart from the traditional pre-occupation with cattle. The authors do justice to the remarkable technical histories of both the diamond and gold mines, and the efforts by government scientists to control serious diseases of livestock, often in the face of resistance by both Boer and black farmers. Few things better illustrate the concept of 'scientific imagination' than the sustained work of Arnold Theiler in veterinary medicine and Hans Merensky in geological prospecting (the Wikipedia entry for the latter reads as follows: *He discovered the rich deposit of alluvial diamonds at Alexander Bay in Namaqualand, vast platinum and chrome reefs at Lydenburg, Rustenburg and Potgietersrus, phosphates and copper at Phalaborwa in the Transvaal lowveld, gold in the Free State, and the world's biggest chrome deposit at Jagdlust near Pietersburg.* Who could do more? How many South Africans know his name?). In the 20th century, figures such as Hendrik van der Bijl and Basil Schonland are deservedly highlighted, simultaneously world-class scientists and inspired institution builders, at the centre of the era called 'The Commonwealth of Knowledge 1930–1948', when Jan Smuts's far-sighted government laid many of the foundations of modern South Africa. The peculiar tendency of independent savantism in a country marginal to the scientific mainstream expressed itself in the globally prescient insights of Eugene Marais, Alex du Toit and Raymond Dart.

In the following period of Afrikaner hegemony and the intensification of apartheid, science and technology were selectively advanced to deal with defence and economic independence, with major developments in chemistry (SASOL, uranium enrichment) and applied nuclear physics. The authors are critical but fair in this section, recognising scientific achievements such as the leadership of yet another superb scientist-administrator, Meiring

Naude, and the invention of the internationally adopted tellurometer by Trevor Wadley and the Helikon vortex tube by Pierre Haarhoff, while deprecating the underlying political motives.

In the last chapter, the authors engage with the democratic 'New South Africa', post-1994. As always in reading a history, it is the present that finds the reader more active in 'comparing notes, as it were' on the lived experience. The authors address the issue of examining indigenous or 'local knowledge' in direct juxtaposition to conventional experimental science; they describe the onset and progression of the HIV epidemic and the denialist response and counter-response; they cover the considerable investment in astronomy and related 'big data'; and provide an excellent review of the work on human origins, both in its older palaeoanthropological and genetic aspects. What is almost completely missing, however, is any mention of the spectacular growth of high-level local expertise in the capacity to study infectious diseases such as HIV infection and TB, where South African groupings in a brace of research universities have earned the respect and support of major international funding organisations for the breadth and depth of their clinical and laboratory contributions. (This now applies also to the current COVID-19 pandemic.) The authors also fall short in their account of the establishment and significant activity of the Academy of Science of South Africa, misattributing the former, short-changing the latter, and failing to see the profound significance of forming a new science academy for all South African scholars, generating evidence-based advice for the whole nation, and embracing consiliently all disciplines which enquire open-endedly after elusive truths by evidence and argument.

Such quibbles do not significantly diminish the value of this otherwise excellent and extremely timely book by two highly seasoned historians who deserve the thanks of all who value the 'scientific imagination', especially here in South Africa but also elsewhere.

AUTHOR:
Remare Ettarh[1]

AFFILIATION:
[1]Health Systems Evaluation and Evidence, Alberta Health Services, Edmonton, Canada

CORRESPONDENCE TO:
Remare Ettarh

EMAIL:
remare.ettarh@albertahealthservices.ca

HOW TO CITE:
Ettarh R. Analysis of citation inequality in Finland and Nigeria using the Lorenz curve. S Afr J Sci. 2021;117(9/10), Art. #10726. https://doi.org/10.17159/sajs.2021/10726

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
citation inequality, Lorenz curve, Gini index, Finland, Nigeria

PUBLISHED:
29 September 2021

Analysis of citation inequality in Finland and Nigeria using the Lorenz curve

Scientometric assessments have become an essential component of research evaluation, and central to these is the use of citation indicators. Citations are used in the performance measurement of research departments and institutions[1], assessment of research funding applications[2], and promotion and hiring of research scientists[3]. Although there have been strong arguments for and against the use of citations for these purposes due to concerns about their reliability and validity as performance indicators, most research institutions across the world continue to make decisions partly based on citation data.[4] Citation counts are still widely used to judge scientific impact and quality, but the significant field-related and temporal variations that exist within the data make comparisons based on absolute counts inappropriate unless done with normalisation.[4]

In the last couple of decades, advances in computational technology have enabled data mining involving citations of millions of scientific papers, resulting in improved understanding of the global citation patterns within and across fields and countries. One emerging pattern of concern is citation inequality, which suggests underlying disparities in research funding, collaboration and the peer-review publication process globally.[5] Non-academic factors such as economic strength, geopolitics, or linguistic and cultural differences, contribute to a global divide in scientific productivity and reward: the wealthier countries which have the top journals and determine which articles get published in them, and the less developed countries with fewer journals and where authors face barriers disseminating their research in the top journals.[6,7] This divide reflects the publishing and citation biases that influence citation inequality within and between countries, regions, and continents. Citation inequality is now evident with respect to individual researchers. A recent study showed that the top 1% most-cited authors accounted for about 20% of citations between 2000 and 2015.[8] It is unclear how patterns of citation inequality differ between specific developed countries and those in Africa. How the distribution of citation shares for individual articles differs between countries is examined here using inequality curves.

Citation data were extracted from Web of Science for all peer-reviewed articles published in all disciplines in 2012 and that had author affiliations in Finland or Nigeria. Lorenz curves were plotted using the cumulative citation shares for all articles from each country. The Gini index, which represents the extent of inequality of article contribution to the citation share based on the Lorenz curve, was calculated for each country. The choice of the countries was partly random, but with a consideration that they were not the top ranked countries on their continents in relation to publications and citations.

The analysis included 11 315 articles from Finland and 2728 articles from Nigeria published between 1 January to 31 December 2012. The choice of year was guided by the greater stability of citations of the articles given the time that has elapsed since then. Finland had a total of 141 627 citations, and Nigeria had a total of 12 600 citations, for the included publications at date of extraction. The Lorenz curves for both countries are shown in Figure 1. The curve for Nigeria shows greater citation inequality (Gini index = 0.59) compared to that for Finland (Gini index = 0.39).

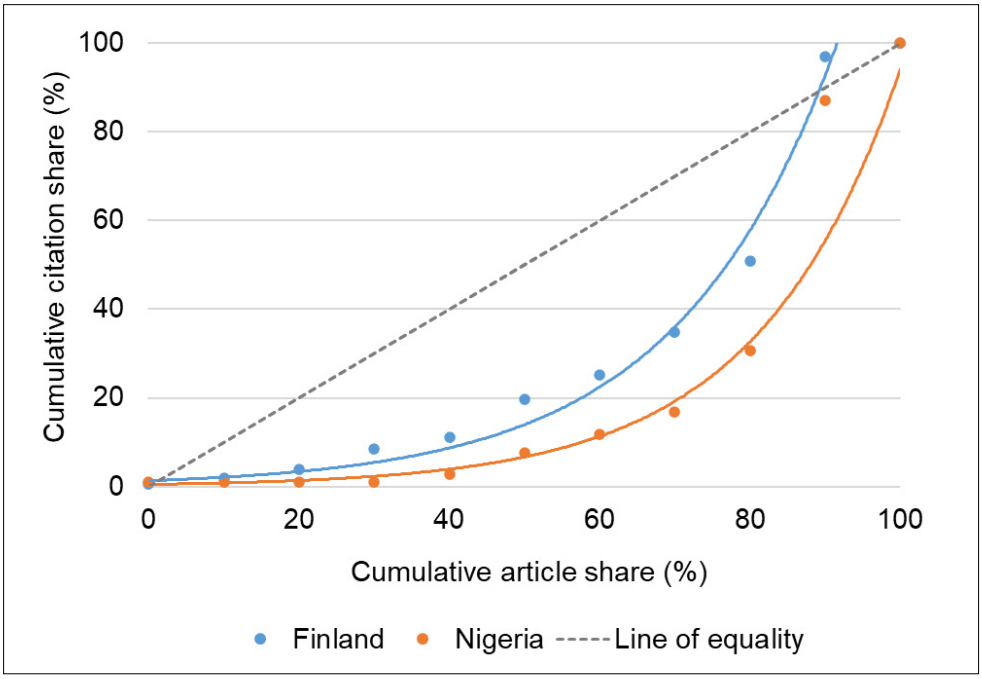


Figure 1: Lorenz curves for cumulative citation and article shares for Finland and Nigeria in 2012.

There are some possible explanations for the difference in the extent of citation inequality between the two countries. Firstly, Nigerian researchers involved in international collaborations and listed as authors of the resulting highly cited papers, such as those from the Global Burden of Disease Study 2010[9], accounted for a large proportion of the citations from the country[10]. Secondly, a significant number of papers was produced from universities in Nigeria but had very few or no citations. This could be attributed in part to the 'Matthew effect', where renowned researchers receive disproportionately greater citations than their lesser-known colleagues for comparable publications[5], in view of the low global visibility of Nigeria-based researchers. This is consistent with a previous observation that Nigerian research communities are largely isolated from partners in other countries inside and outside Africa.[11] Interestingly, international research collaborations have been increasing over the past century across various disciplines.[12] In addition, the results reflect a more equitable spread of the bibliometric impact of research articles from Finland.

The significance of this analysis is in the demonstration of the use of the Lorenz curve and Gini index to compare citation inequality between countries, and the possibilities it creates to examine these patterns across other African countries and periods of time. This approach reveals the extent of the influence of highly cited papers on the citation inequality observed in each country, and the weight of scientific knowledge that is largely ignored or unused as seen in the long tail of the curve. This type of analysis could provide the relevant government departments in these countries with a snapshot of the performance of the domestic research community, and help in assessing the impact of strategies to improve international collaboration and enhance the use of the knowledge produced by in-country researchers.

References

1. Aroeira RI, Castanho MARB. Can citation metrics predict the true impact of scientific papers? FEBS J. 2020;287(12):2440–2448. https://doi.org/10.1111/febs.15255
2. Yan E, Wu C, Song M. The funding factor: A cross-disciplinary examination of the association between research funding and citation impact. Scientometrics. 2018;115:369–384. https://doi.org/10.1007/s11192-017-2583-8
3. Moher D, Naudet F, Cristea IA, Miedema F, Ioannidis JPA, Goodman SN. Assessing scientists for hiring, promotion, and tenure. PLoS Biol. 2018;16(3), e2004089. https://doi.org/10.1371/journal.pbio.2004089
4. Aksnes DW, Langfeldt L, Wouters P. Citations, citation indicators, and research quality: An overview of basic concepts and theories. SAGE Open. 2019;9(1):1–17. https://doi.org/10.1177/2158244019829575
5. Xie Y. "Undemocracy": Inequalities in science. Science. 2014;344(6186):809–810. https://doi.org/10.1126/science.1252743
6. Demeter M. The winner takes it all: International inequality in communication and media studies today. J Mass Commun Q. 2019;96(1):37–59. https://doi.org/10.1177/1077699018792270
7. Zanotto SR, Haeffner C, Guimaraes JA. Unbalanced international collaboration affects adversely the usefulness of countries' scientific output as well as their technological and social impact. Scientometrics. 2016;109:1789–1814. https://doi.org/10.1007/s11192-016-2126-8
8. Nielsen MW, Andersen JP. Global citation inequality is on the rise. Proc Natl Acad Sci USA. 2021;118, e2012208118. https://doi.org/10.1073/pnas.2012208118
9. Lim SS, Vos T, Flaxman AD, Danaei G, Shibuya K, Adair-Rohani H, et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: A systematic analysis for the Global Burden of Disease Study 2010. Lancet. 2012;380(9859):2224–2260. https://doi.org/10.1016/S0140-6736(12)61766-8
10. Hsiehchen D, Espinoza M, Hsieh A. Multinational teams and diseconomies of scale in collaborative research. Sci Adv. 2015;1(8), e1500211. https://doi.org/10.1126/sciadv.1500211
11. Ettarh R. Patterns of international collaboration in cardiovascular research in sub-Saharan Africa. Cardiovasc J Afr. 2016;27(3):194–200. https://doi.org/10.5830/CVJA-2015-082
12. Larivière V, Gingras Y, Sugimoto CR, Tsou A. Team size matters: Collaboration and scientific impact since 1900. J Assoc Inf Sci Technol. 2015;66:1323–1332. https://doi.org/10.1002/asi.23266

AUTHORS:
Farai Nyabadza[1,2]
Alex Broadbent[3]
Charis Harley[2,4]
Abejide Ade-Ibijola[2,5]
Ebrahim Momoniat[1,2]

AFFILIATIONS:
[1]Department of Mathematics and Applied Mathematics, University of Johannesburg, Johannesburg, South Africa
[2]Data Science Across Disciplines Research Group, Institute for the Future of Knowledge, University of Johannesburg, Johannesburg, South Africa
[3]Institute for the Future of Knowledge, University of Johannesburg, Johannesburg, South Africa
[4]Faculty of Engineering and the Built Environment, University of Johannesburg, Johannesburg, South Africa
[5]Department of Applied Information Systems, College of Business and Economics, University of Johannesburg, Johannesburg, South Africa

CORRESPONDENCE TO:
Ebrahim Momoniat

EMAIL:
emomoniat@uj.ac.za

HOW TO CITE:
Nyabadza F, Broadbent A, Harley C, Ade-Ibijola A, Momoniat E. Models and muddles in the COVID-19 pandemic. S Afr J Sci. 2021;117(9/10), Art. #9506. https://doi.org/10.17159/sajs.2021/9506

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
mathematical models, epidemiology, infectious diseases, public health, COVID-19

PUBLISHED:
29 September 2021

Models and muddles in the COVID-19 pandemic

> *It is a capital mistake to theorise before one has data. Insensibly one begins to twist facts to suit theories, instead of theories to suit facts.*
>
> Sir Arthur Conan Doyle

The COVID-19 crisis is an opportunity for scientists to showcase their skill and the impact that good science can have on society. However, not all scientists have risen to the occasion with the sense of responsibility and accountability that their work deserves. Scientists worldwide have shown, and continue to show, great enthusiasm regarding the use of specific scientific tools, mainly modelling and predictive analytics, to estimate how the virus spreads and behaves and to assess interventions against counterfactual scenarios. In this Commentary, we question whether the application of these tools has always been appropriately managed by discussing the underlying elements of modelling which need to be understood and evaluated for results to be meaningful and credible.

A mathematical model must capture the principles that dictate the dynamics of what is being modelled: assumptions, constraints and relevant natural laws, for example. These principles serve as the 'rules' for understanding the results obtained and provide the context within which the model has meaning. Within this context, the model goes beyond being a mere collection of mathematical operations and represents – albeit in idealised or imperfect form – some feature of the actual world. Here we argue that these rules have often been ignored when engaging with the results obtained from mathematical models used for predictive purposes in the COVID-19 pandemic (including policy purposes) and from data-driven models designed via machine-learning methods.

To make our case, we first provide some context of the origins of disease modelling and then offer a 'current day' frame of reference which illustrates why caution is needed when employing models for prediction.

Infectious disease modelling

Infectious disease modelling is one small part of infectious disease epidemiology, which is a small part of epidemiology. How, then, did modelling come to dominate, not only the prediction of the spread of COVID-19, but also policy decisions with consequences reaching far beyond the death toll of the disease itself?

It is helpful to understand some of the conceptual evolution of epidemiology from its foundations, which were laid during the era of industrialisation in Europe.[1] Cities grew, bringing people into close proximity, many of them malnourished, with poor or non-existent sanitation and hygiene practices. Infectious disease flourished in these newly swollen human ecosystems, and epidemics were a regular occurrence. At the same time, information on disease incidence and deaths became readily available in concentrated form for the first time. Enterprising medical thinkers realised they might infer the causes of outbreaks from this information. Thus two new ways of thinking were born together, with epidemics lending their name to one epidemiology and the state that collected information lending its name to statistics.

It was epidemiology that taught us the health significance of personal hygiene. Epidemiology helped us uncover the 'germ theory' of disease[2], which ultimately turned out to account for infectious diseases and is central to contemporary Western medicine development.

Germ theory and hygiene theory were at odds during this period, with 'hygienists' seeing the germ theory as an attempt to let the authorities off the hook of ensuring more humane living conditions for the labouring classes.[3] The germ theory 'won' in the theoretical sense. The 'miasma' theory of disease was ultimately discredited: diseases are not caused by bad smells, and vaccination proved very effective. However, the recommendations of the hygienists were effective in another sense, because personal hygiene and good sanitation are necessary for people to dwell together in city conditions even with the technology of vaccination.

Fast forward to the mid-20th century, and the advent of antibiotics and other innovations such as ventilation put many infectious diseases even further on the back foot, even though viruses remained stubbornly resistant to direct medical solution. In the 1940s and 1950s, epidemiological attention shifted to another 'epidemic': that of lung cancer, which had rocketed from a virtually unknown condition during the previous two decades. Why?

Through a remarkable methodological development, Sir Austin Bradford Hill and others perceived the significance of an apparently tiny difference between the odds of smoking among people living with lung cancer and among people living with other cancers, which translated into a remarkably large risk ratio.[1] They were able to anticipate and check for an extensive range of plausible confounders. With others, such as Jerome Cornfield[4], they were able to corral evidence from other domains for a causal connection between smoking and lung cancer and against the leading rival hypothesis: the 'constitutional hypothesis' that some gene caused both.

In this episode, the modern discipline of epidemiology was born. Two characteristics are relevant to the present narrative. Firstly, this was a discipline of campaigning. The real 'win' for these epidemiologists was not the scientific case that smoking caused lung cancer, but the recognition in 1964 in the US Surgeon General's report that smoking causes lung cancer.[5] This led to a series of regulations and public health advice – changes that were fought by tobacco companies and continue to be fought in many Asian countries today.[6]

James Lind had sought to convince the naval authorities to provide limes to sailors against scurvy, eventually succeeding (hence the term 'limeys' for British sailors). Less successfully, the problematic and abrasive Ignaas Semmelweis had sought to instigate hand washing in the General Hospital in Vienna. In London, John Snow persuaded authorities to remove the water pump handle in Broad Street, which was drawing water from the polluted River Thames to create an epicentre of cholera in the district.[1]

But with the smoking and lung cancer episode, the political engagement of epidemiology was affirmed for chronic diseases. Contemporary epidemiologists continue not only to seek the scientific truth about, for example, sugar and all-cause mortality but also to campaign for sugar taxes.

The second feature of the Bradford Hill story is its informality, immortalised in nine 'viewpoints' for assessing causality.[7] Bradford Hill urged epidemiologists to consider causal hypotheses from various perspectives. How strong is the association? Is a causal link biologically plausible? Does evidence from basic sciences support it? Is there a dose-response relationship? And so forth. These were subsequently interpreted as a checklist and remain in use today. But that was never the intention. They were not meant as *sine qua non* for causal inference but as guides to the ultimate question: Is there any hypothesis that better explains the evidence than that of cause and effect?

While epidemiology subsequently developed in mathematical complexity, this relatively informal, subjective approach did not. Causal inference remains a stubborn philosophical problem, and so the inability to define and proceduralise it is not surprising. But it is something of a challenge when peers seek to assess each other's work. It is also hard to teach and something of an embarrassment for those who prefer to think of epidemiology as closer to the natural than the social sciences. Efforts to formalise causal inference now form a considerable part of the epidemiological methodological agenda, and are a growth area for the discipline.[8,9]

In infectious disease epidemiology, modelling provides a way to formalise the central question, which concerns predicting the course of a disease, even after the cause is known (as is more often the case now, where it was not a hundred years ago). Modelling makes use of new computing power and enables the consequences of assumptions to be worked out in detail. This highlights the nature and justification of the assumptions themselves, the sensitivity of predictions to those assumptions and inaccuracies in the data – all of which is beneficial. It also enables predictions that are much better supported than would otherwise be the case – provided the assumptions themselves are well supported and the data reliable.

Thus, contemporary epidemiology is influenced by two paradigm-shaping instincts: the sense that campaigning for public health policy is part of the epidemiological mandate and the desire for methodological progression within the science towards more formal approaches.

Models in the making

All models are wrong, but some are useful.

George E. P. Box

During the COVID-19 pandemic, the tradition of campaigning and the associated sense of urgency may have contributed towards some unfortunate lapses in the use of models. Models are abstract representations of real phenomena, and are useful for making predictions. At best, a good model has two facets: accuracy and simplicity. The accuracy is vital in linking the model to reality, while simplicity is paramount for understanding. Despite their usefulness, models are always shrouded with limitations. We discuss some of these in the present section.

Two important considerations – assumptions and reliable data – were often not transparently communicated or verified, which naturally had an impact on the effectiveness of South Africa's response which mathematical models primarily influenced. A significant number of these models has been published globally since the beginning of the pandemic.[10,11] Given that there is no known effective pharmaceutical treatment for COVID-19 (at the time of writing), mathematical models have shaped policy with respect to non-pharmaceutical interventions, intending to limit transmission between persons and contaminated environments, and in so doing 'flatten the curve' of infected persons.

Lessons learnt from the SARS outbreak in 2003 and the MERS outbreak in 2002 provided a medical understanding of how coronaviruses affect the lower respiratory tract. Taiwan was one of the first countries to implement the non-pharmaceutical interventions learnt from SARS. Some of these strategies include the wearing of masks and contact tracing. Modelling helps us to quantify the impact of these preventative measures on the spread of the disease.

From a modelling perspective, the integrity of a mathematical model is in its assumptions, consanguineousness to the available data, and the power to predict the epidemic trends in the short or long term. In building mathematical models for the pandemic in South Africa, it is crucial to consider the following: heterogeneity in the population densities, economic realities, inconsistent policies and inharmonious enforcement of regulations. Model-building is thus not an abstract exercise but requires deep contextualised knowledge. In South Africa, appreciation of socio-economic dynamics is critical in the modelling process, such as overcrowding, a large informal sector, and high levels of poverty. Within this context, social distancing in South Africa, for instance, should be understood within the context of a heterogeneous distribution of populations and varied patterns of movement within and between cities and provinces. These social realities have implications for the pandemic's spread as densely populated areas, such as public transport hubs and spaces, are hotspots for transmission. The resurgence of the epidemic in many countries has seen the emergence of a more recent dynamic termed a 'superspreader event', threatening the fragile equilibrium South Africa has achieved. Furthermore, as prevention fatigue sets in, the relaxation of preventative efforts can be a source of disease recrudescence. Models that capture such scenarios are of interest from a policy formulation and disease management perspective.

There is always a trade-off between model complexity and its tractability – the more complex the model, the less tractable and vice-versa. Many of the recent models had a few noticeable challenges when it came to functioning as workable solutions. Firstly, the overestimation of predicted numbers led to panic amongst the public, given a poor understanding among most that, while models are useful tools, they should not be over-interpreted, especially when considering long-term projections. Being dynamic, the implementation of an intervention of any kind will immediately impact the progression and trajectory of the disease described by the model. Secondly, many models have been built to provide predictions for scenario planning without clearly explaining the underlying assumptions which inform these predictions. Lastly, models depend on assumptions, and the sensitivity to errors in these assumptions should be aligned to the social-economic dynamics of a given setting to create realistic outcomes.

Given the complexity of a functioning society with varied dynamics, models should ideally be interdisciplinary. The role of social, cultural and human behaviour and economic consequences of the pandemic and any possible interventions cannot be ignored when modelling a pandemic. Thus, while the role of mathematical models as tools for understanding the transmission dynamics of COVID-19 in South Africa cannot be underestimated, one thing needs to be kept in mind: all models are necessarily approximations of the real world, being simplifications of reality driven by the need to answer specific questions and in many cases one particular question. Models capture certain aspects of a phenomenon, under certain assumptions, while relying on relevant data sets where the quality, accuracy, specificity, availability and usability of the data are key to the usefulness of a model.

The big difference in 2020/2021 has been the impact of social media and greater transparency, which meant that modelling as a tool took a leading role in combatting COVID-19. Scientists involved in the modelling of MERS/SARS did not benefit from exposure to social

media. During previous disease outbreaks, only mathematicians had access to resources to do the modelling, and often the products of such engagement were kept within the academic community. The power of modern computers allows anyone with basic knowledge to develop models of the spread of COVID-19. The effective use of social media provides for wide dispersal of the forecasts and information learnt from such models.

Data-driven models

There have been ongoing concerns about the quality and availability of data relating to the current pandemic.[12] Yet, aside from being employed for the development of predictive models, these questionable data are underlying some of the most critical metrics that are being used to gauge our progress in fighting the novel coronavirus. Hugely debated results published in a preprint in April 2020 present two antibody tests conducted by universities in California. It was claimed that possibly 28 to 85 times more people had been exposed to COVID-19 than had been detected using the PCR method.[11] If so, a slight alteration in the relevant denominator used to calculate the mortality rate from COVID-19 would have shifted mortality from a figure of 2.5–3%, which public health officials had been working with, to between 0.12% and 0.20%. Given that seasonal influenza's mortality rate is about 0.1%, it is clear that such a change in the mortality rate describes an entirely different pandemic.[13]

Recent events seem to support the argument that we may be overestimating the infection fatality rate of COVID-19. In September 2020, it was indicated that upon taking into account asymptomatic cases, the infection fatality rate had shifted from between 2% and 3% to 1%.[14] More recent updates by the CDC and WHO indicate an infection mortality rate of between 0.65% and 0.5–1%, respectively, while the work by Ioannidis[15] indicates a median infection fatality ratio of 0.23% across 51 locations. The conclusion to be reached is that when there are errors in raw data or our data sets are limited and constantly changing, we need to be cautious about reaching conclusions regarding the nature of a phenomenon. Inappropriate or inaccurate statistics can generate an image that differs vastly from the reality they are trying to capture.

It is no secret that, similar to mathematical model development, data science relies heavily on assumptions made in the scientific process. One of these assumptions is that the data set used in the scientific process represents the studied entity/population. Hence, when data are skewed, limited, or contextually inappropriate, the results of the whole scientific process are most likely to be incorrect. The idea of 'wrong data' has been identified as artificial intelligence's (or AI's) biggest risk factor.[16]

There are concerns regarding the quality of data from all countries, often for similar reasons.[12] Instead of national health surveillance systems that can be relied upon to provide reasonably accurate data, there is a patchwork of voluntary data-gathering processes in place at most hospitals. Naturally, not all hospitals report the data, and the data are not consistent from hospital to hospital.[12] Furthermore, delays in obtaining data from hospitals and other health facilities lead to data that do not reflect the current situation on the ground.[17] When models employ such data for predictive purposes, they end up *predicting the past*. The use of contextually inappropriate data to design a model is also concerning. An example of this is the use of data from China to predict the spread of South Africa's epidemic. Where machine learning has been used with inappropriate data, the results that have been obtained are misleading. Yet the implementation and reporting of results obtained from machine-learning models are on the increase. Why is that?

Conclusion

Epidemiology has a long and fine tradition of engaging public policy to change it. The discipline also naturally seeks technical development of its methodology. In the case of COVID-19, these two instincts – campaigning and formalising – came together in an unfortunately unholy alliance. As has been remarked, infectious disease modelling is only one part of infectious disease epidemiology; but it is particularly striking because it appears to represent methodological progression. When the campaigning instinct kicks in, there is a danger of overreach. Policies are pushed that simply fail to consider all factors because the models do not consider all factors. Both epidemiology and public health can do better in the future by considering a more extensive range of health consequences beyond the deaths of people with the virus in their bloodstream, by being less belligerent about the importance of these consequences above other policy priorities. Humility is an epistemic virtue.

As current data on the coronavirus are not reliable in the sense that we are constantly adjusting for inaccuracies or obtaining new information, our aim should rather be to create opportunities for further research on the novel coronavirus in the future. As such, our goal should be to create data repositories, structure data cleaning processes, design data pipelines, and develop better tools to model past pandemics to gain a deeper understanding for further comparison. Learning from the past should be our strategy to become a society that reflects on past mistakes, assesses current inadequacies, and then moves forward with greater awareness and humility.

Acknowledgements

E.M. acknowledges support from the National Research Foundation of South Africa (grant no. 103483).

Competing interests

We have no competing interests to declare.

References

1. Morabia A. History of epidemiologic methods and concepts. Basel: Birkhauser Verlag; 2004. https://doi.org/10.1007/978-3-0348-7603-2
2. Broadbent A. Why philosophy of epidemiology? Philosophy of Epidemiology. London: Palgrave Macmillan; 2013. https://doi.org/10.1057/9781137315601
3. Carter KC. The rise of causal concepts of disease: Case histories. London: Routledge; 2003. https://doi.org/10.4324/9781315237305
4. Cornfield J, Haenszel W, Hammond EC, Lilienfeld AM, Shimkin MB, Wynder EL. Smoking and lung cancer: Recent evidence and a discussion of some questions. Int J Epidemiol. 2009;38:1175–1191. https://doi.org/10.1093/ije/dyp289
5. United States Surgeon General's Advisory Committee on Smoking. Smoking and health: Report of the Advisory Committee to the Surgeon General of the Public Health Service [webpage on the Internet]. c1964 [cited 2021 Aug 07]. Available from: https://profiles.nlm.nih.gov/101584932X202
6. Broadbent A, Hwang S-S. Tobacco and epidemiology in Korea: Old tricks, new answers? J Epidemiol Community Health. 2016;70:527–528. https://doi.org/10.1136/jech-2015-206567
7. Bradford Hill A. The environment and disease: Association or causation? Proc R Soc Med. 1965;58:259–300. https://doi.org/10.1177/003591576505800503
8. Rubin DB. Causal inference. In: Tierney R, Peterson P, Baker E, McGaw B, editors. International encyclopedia of education. 3rd ed. Amsterdam: Elsevier; 2010. p. 66–71. https://doi.org/10.1016/B978-0-08-044894-7.01313-0
9. Van der Weele TJ. Explanation in causal inference: Methods for mediation and interaction. Oxford: Oxford University Press; 2015.
10. Enserink M, Kupferschmidt K. Mathematics of life and death: How disease models shape national shutdowns and other pandemic policies. Science Insider. 2020 March 25; Health. https://doi.org/10.1126/science.abb8814
11. Chowdhury R, Heng K, Shawon MSR, Goh G, Okonofua D, Ochoa-Rosales C, et al. Dynamic interventions to control COVID-19 pandemic: A multivariate prediction modelling study comparing 16 worldwide countries. Eur J Epidemiol. 2020;35:389–399. https://doi.org/10.1007/s10654-020-00649-w
12. Irfan U. How common is Covid-19? What 2 controversial antibody studies can and can't tell us. Vox. 2020 April 24 [cited 2021 Aug 07]. Available from: https://www.vox.com/2020/4/24/21229415/coronavirus-antibody-testing-covid-19-california-surve
13. Woodie A. How the lack of good data is hampering the COVID-19 response. Datanami. 2020 April 27 [cited 2021 Aug 07]. Available from: https://www.datanami.com/2020/04/27/how-the-lack-of-good-data-is-hampering-the-covid-19-response/

14. Yong SJ. Clarifying the true fatality rate of Covid-19: Same as the flu? Medium. 2020 September 08 [cited 2021 Aug 07]. Available from: https://medium.com/microbial-instincts/clarifying-the-true-fatality-rate-of-covid-19-same-as-the-flu-8148e38b9ab5

15. Ioannidis JPA. Infection fatality rate of COVID-19 inferred from seroprevalence data. Bull WHO. 2020 October 14. https://doi.org/10.2471/BLT.20.265892

16. Korolov M. A.I.'s biggest risk factor: Data gone wrong. InsiderPro. 2018 February 13 [cited 2021 Aug 07]. Available from: https://www.idginsiderpro.com/article/3254693/ais-biggest-risk-factor-data-gone-wrong.html

17. Short A. Reporting delays and political pressure contribute to the CDC's incomplete picture of COVID-19. Business Insider. 2020 May 19 [cited 2021 Aug 07]. Available from: https://www.businessinsider.com/why-the-cdc-coronavirus-data-is-so-bad-2020-5?IR=T

AUTHORS:
Lucas Hertzog[1]
Camille Wittesaele[2,3]
Raylene Titus[1]
Jenny J. Chen[3]
Jane Kelly[1]
Nontokozo Langwenya[3]
Lauren Baerecke[1]
Elona Toska[1,3,4]

AFFILIATIONS:
[1]Centre for Social Science Research, University of Cape Town, Cape Town, South Africa
[2]Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, United Kingdom
[3]Department of Social Policy and Intervention, Oxford University, Oxford, United Kingdom
[4]Department of Sociology, University of Cape Town, Cape Town, South Africa

CORRESPONDENCE TO:
Lucas Hertzog

EMAIL:
lucas.hertzog@uct.ac.za

HOW TO CITE:
Hertzog L, Wittesaele C, Titus R, Chen JJ, Kelly J, Langwenya N, et al. Seven essential instruments for POPIA compliance in research involving children and adolescents in South Africa. S Afr J Sci. 2021;117(9/10), Art. #12290. https://doi.org/10.17159/sajs.2021/12290

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
POPIA, children and adolescents, protection of personal information, data management, research governance

PUBLISHED:
29 September 2021

Seven essential instruments for POPIA compliance in research involving children and adolescents in South Africa

Compliance with South Africa's *Protection of Personal Information Act* (POPIA) is a foremost governance challenge for research involving high-risk and vulnerable groups such as children and adolescents. It remains unclear what constitutes adequate safeguards to protect the personal information of the child under this new law. To meaningfully adhere to the principal aims of POPIA, researchers must understand and address the implications of this legislation on research governance practices. Navigating the additional POPIA compliance requirements within established research projects additionally raises questions about how research can use POPIA to build on existing research governance mechanisms without extreme additional burden on research teams.

We invite readers to explore a series of best practices in safeguarding the personal information of children, adolescents, and young people (0–24 years old) – a key age group that represents nearly half of South Africa's population in 2021. We will discuss possible actions which can be taken to ensure POPIA effectively builds on existing data protection mechanisms for research projects at all stages of the research cycle. These actions promote compliance to POPIA throughout the data life cycle. Our objective is to stimulate a broader conversation on how to improve the protection of children's and adolescents' sensitive personal information in South Africa and inform considerations that need to be addressed by the POPIA Research Code of Conduct.

We join the POPIA discussion as a research group generating evidence that influences social and health policy and programming for young people in sub-Saharan Africa. Our contribution draws on our work adhering to multiple transnational governance frameworks imposed by national legislation such as data protection regulations, funders, and academic institutions. This has involved the use of several research governance mechanisms. In this Commentary, we summarise seven essential instruments to assist research projects involving children and adolescents to achieve POPIA compliance.

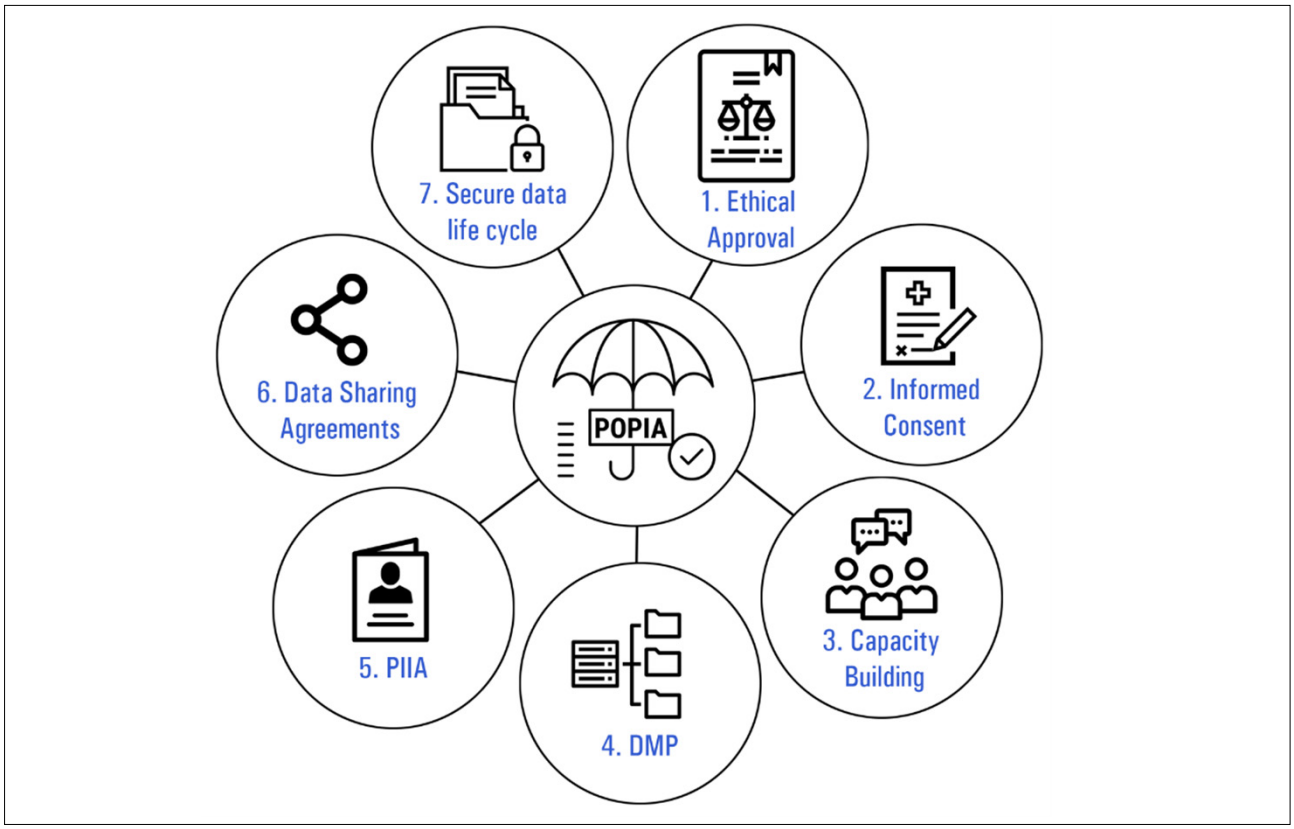
POPIA and research with children and adolescents

Since July 2021, researchers have been mandated to comply with South Africa's newly ratified *Protection of Personal Information Act* (POPIA), *No. 4 of 2013.* This Act has implications for all research involving vulnerable populations such as children and adolescents. Compliance mechanisms must be adapted to secure children's and adolescent's rights to privacy while balancing other rights and interests such as participating in research of public interest. Under POPIA, the ban on processing personal information relating to a child has research-specific exceptions that clarify the terms under which children's data can be processed. This is stated in Section 35 alongside regulations on prior consent of a competent person for data collection, with specific provisions in Section 11. Nevertheless, the scientific community is uncertain about which specific mechanisms to implement towards POPIA compliance within their research projects and what concrete changes are needed to research governance structures and processes. POPIA triggers additional complexity for transnational research collaborations requiring a reconciliation with other data protection regulations such as the European Union's General Data Protection Regulation (GDPR). POPIA mentions that 'appropriate' and 'adequate' safeguards should be implemented to protect the personal information of minors. It remains unclear what constitutes appropriate and adequate for protecting the personal information of children, adolescents and young people (0–24 years old)[1] who represent nearly half of the South African population[2].

As researchers, it is our ethical obligation to safeguard the rights and interests of research participants. We must also comply with national legislation, regulations and Codes of Conduct imposed by governments, research institutions, and funders. The roll-out of POPIA has created an opportunity to implement improved safeguards for the secure processing of personal information in our research. We have been adjusting our research governance, ethics, and data management processes to meet regulatory frameworks across multiple countries, research institutions, and studies. This has resulted in the development and use of seven essential instruments that are aligned with our vision to generate rigorous evidence for the public interest.

Stimulated by recent discussions initiated by the Academy of Science of South Africa (ASSAf) and the development of the POPIA Code of Conduct for Research[3], we propose instruments to support POPIA compliance in the context of research with children and adolescents. Our objective is to: (1) share our research group's experience in safeguarding the protection of children and adolescent's personal information; (2) outline the seven essential instruments employed by our team to comply with POPIA and GDPR (Figure 1); and (3) stimulate a discussion on how to improve the protection of sensitive personal information within research contexts in South Africa.

The following research governance and data management instruments were constructed in the context of our research consortium primarily located in South Africa and the UK. Not limited to, but primarily focusing on data collected in South Africa, our vision is to generate scientifically rigorous evidence to influence policy and programmes to support children and adolescents to reach their full potential. Supported by various funders, we work in close collaboration with policy and programming stakeholders such as UNICEF, WHO, UNDP, the South African and other African governments. These instruments were constructed to safeguard personal information in longitudinal social science studies and randomised trials, including a large cohort of adolescents living with HIV, a cohort of adolescent mothers and their children, and several studies of parenting programmes in low- and middle-income countries.



DMP, data management plan; PIIA, personal information impact assessment

Figure 1: Seven essential instruments for POPIA compliance in multi-partner research among children and adolescents.

Instrument 1: Enhanced ethical approval from research ethics committees

Ethical responsibility to our research participants is our foremost concern. Nevertheless, ethical clearance processes from existing research ethics committees may no longer be sufficient in the era of increased digitalisation. Research ethics committees are demanding increasingly detailed information about the handling of personal information and in some cases will make favourable ethical approvals contingent on the opinion of an information regulator such as an institution's information officer, who may or may not be involved in research ethics committees directly. Acquiring clearance for processing personal information from recognised authorities should be regarded as equally important (in terms of timelines, resourcing, and compliance) to acquiring ethical clearance from existing research ethics committees. Each must be held in continuous review and monitored simultaneously.

Section 34 of POPIA prohibits the processing of personal information of minors unless provisions of Section 35 are applicable. To meet these special provisions, ethical approvals from institutional review boards are key to confirming whether the research and processing of personal information are appropriate and for the public interest. Some ethical parameters may be set out in research ethics applications enabling researchers to ensure personal information provided by data subjects is protected in accordance with principles of ethical research and protection of personal information. Our ethical and methodological procedures are informed by over 12 years of fieldwork experience in multiple South African provinces[4] and other studies working with vulnerable populations in comparable contexts[5,6], and are summarised below:

1. Obtain ethical clearance from suitable research ethics committees.
2. Obtain consent of each individual involved at each stage of data collection.
    a. Consent must be provided by a competent person where data subject is a minor, followed by participant assent.
    b. Consent forms must clearly identify institutions and lead investigators responsible for data management: cleaning, analyses and sharing.
3. Maintain data subject's confidentiality in line with consent form.
4. De-identify data sets at earliest opportunity and minimise the risk of re-identification.
5. Limit access to personally identifiable information on a 'need-to-know' basis.
6. Appropriate retention of personal information records for historical, statistical, and research purposes with sufficient safeguards against the records being used for unauthorised purposes.
7. Responsible parties should take measures to ensure that personal information is always secure throughout data collection, processing, migration, storage, sharing, archiving, and dissemination.
8. Mechanism for personal information to be withdrawn at request by the data subject and competent person.
9. Data subjects are interviewed by trained interviewers in private locations to maximise confidentiality.

Instrument 2: Informed consent for use of personal information and voluntary participation

The process of obtaining informed and voluntary consent from research participants is central to conducting ethical research.[7] In basic terms, it aims to ensure individuals are adequately informed about the risks and benefits prior to providing voluntary consent for participation. In the context of research with children and adolescents under the age of 18, consent is also a process of dialogue with caregivers in respect

of their child's rights. With the roll-out of data protection regulations, researchers should also use this process to inform children and their caregivers about their rights regarding their personal information and privacy.[8] This renders the informed consent process a fundamental instrument for enabling data subjects to be informed about the risks and benefits of providing personal information prior to providing voluntary consent.

When working with vulnerable groups, such as children and adolescents, power relations and additional considerations should be accounted for. Informed consent, and assent in the case of those under 18, should be obtained from data subjects and a competent person prior to data collection. When approaching caregivers whose children might be eligible for a study, it must be clear that they are a competent person who can allow their child's personal information to be processed by a data operator and responsible party. Data collectors must be cognisant about the inherent power discrepancy at play throughout the data life cycle. This is particularly important in South Africa, where adolescents and caregivers may have low literacy rates.[9] To mitigate the effect of low literacy rates, information sheets and consent forms should be constructed using accessible language. Both documents should always be read aloud to the data subjects and competent person in their chosen language. Ample opportunities should be offered for the data subject and competent person to ask questions and decide about participation. Continuous consent should be obtained by requesting consent prior to each research activity and phase. If a participant withdraws consent or requests for their data to be withdrawn, the data operator should regain consent for the processing of any personal information that was previously collected.

POPIA places emphasis on the 'specific' expression of informed consent by the data subjects. Researchers should ensure that data subjects are informed about: (1) which data protection regulations govern the handling of personal information in the research, (2) the nature of data that will be collected, (3) how it will be processed, (4) where it will be stored, (5) what security measures will be in place to protect the data, (5) who will have access to personal information, (6) how long their data will be retained, and (7) how a data subject may request for their data to be updated or removed. Particularly in longitudinal studies, researchers should ensure explicit permission from data subjects is obtained to contact them in the future.

Instrument 3: Capacity building and knowledge co-creation

POPIA triggers the need for capacity sharing spaces for researchers and potentially research participants. Such spaces empower individuals to learn about their rights and equip researchers to obtain essential research skills. It also facilitates opportunities to enter discussions with POPIA experts about the implications on research. Several opportunities were mechanised in our team to simultaneously build understanding about the provisions of POPIA while moving towards compliance throughout the studies within our research group.

First, an internal forum with experts was established to provide capacity-building for our team. These focused on topics such as: special categories of data, specific provisions for research, data sharing, and demonstrating POPIA compliance to research ethics committees and information regulators. These training sessions were instrumental to enhance understanding of the Act and its implications on research and stimulated an internal POPIA-informed audit in our research governance documents and protocols.

Second, to assess what adaptations each study should implement to ensure POPIA compliance, a research governance team was established within our research group. Study leads of each study completed risk assessments in their data management plans for the research governance team to identify the type and format of personal information collected within each study, including high-risk information.[3]

Finally, the research governance team proceeded to identify training needs of data operators and data collectors. The team implemented several data security enhancement processes such as direct uploading of data onto protected servers using end-to-end encryption instead of password-protected laptops. However, being POPIA compliant during remote data collection – due to COVID-19 safety requirements – requires reflections on: (1) how researchers provide participants with a copy of the informed consent forms when working in resource-limited settings, (2) how to maintain and track the process of consent (verbally, text messages and voice recordings) for each phase of data collection, including when providing referrals, and (3) ensuring participants have contact details for the research team, ethics committees and responsible parties. Importantly, it is critical to reflect how one can ensure confidentiality of sharing informed consent forms given the high rates of sharing of mobile devices in South African homes.[10] Compliance requires strengthening existing consent processes when sharing a copy of the informed consent forms via mobile text, deciding on safe mechanisms to share consent audio recordings with the participant without breaching their privacy and confidentiality. These considerations introduce additional administrative tasks for data collectors who need ongoing support in transition to meeting POPIA's requirements.

Instrument 4: Data management plan

To ensure each study within the research consortium has ownership of their data management practices, each study maintains: (1) a data management plan and (2) personal information impact assessments. This ensures that each document is tailored to the unique requirements of each study within the research group. The development of a data management plan assists study teams to make decisions about how research data will be handled throughout the data life cycle (i.e. collection, processing, analysing, preserving, sharing, and archiving). Carefully planning and agreeing on how data will be managed at the outset, and keeping this in review, minimises data protection risks and enhances the public benefit of research. This document should be brief to promote its use and adherence and cover at least four key elements (Table 1). Data management plans should be treated as living documents, to be maintained throughout the data life cycle, triggered by key research cycle events: (1) when substantive changes in data needs arise, (2) at scheduled timepoints, and/or (3) at key study stages.

Table 1: Four key elements of a data management plan

Element	Indicative questions
1. Data collection and description	Will you produce original data and/or use existing data? Where and how will you get your data? What types and format of data will you collect and how will you describe them?
2. Data curation and storage	Where will you store your data? How will you organise and name your data files? How large are your data?
3. Data security	What provisions for secure storage and transfer of sensitive data are in place? Are the data safely stored in repositories for long-term preservation?
4. Data sharing and reuse	How and where will you share your data during and after the study? What are your plans for long-term data sharing and preservation? Who will have access, under what conditions and for how long?

Instrument 5: Personal information impact assessment

Conducting a personal information impact assessment (PIIA) is an instrumental process for evaluating compliance with POPIA when processing special category data. Unlike GDPR, POPIA does not contain equivalent provisions for data protection impact assessments. However, POPIA does outline that an information officer may conduct a PIIA to evaluate whether adequate measures are enacted to comply with POPIA. Given that our research involves a vulnerable population, each study maintains both a data protection impact assessment and PIIA to ensure that safe and lawful processing of personal information are embedded in each study by design. Importantly, this enables researchers to demonstrate and ascertain effective compliance with each regulation. It is advised that emphasis be placed on the rights and interests of the data subjects whose personal information is being processed when completing these assessments. PIIAs call on the responsible party to consider the necessity and proportionality for processing personal information. It also includes a risk assessment detailing the potential risks to data subjects, so the effectiveness of risk mitigation measures can be reviewed. Each study also has a personal data workflow which captures and presents the flow of all personal information throughout the life cycle of the project (Figure 2). As with data protection impact assessments, this assessment should then be reviewed by a relevant authority. In tandem with the data management plan, PIIAs should be maintained as living documents.

Instrument 6: Collaboration and data sharing agreements

Successful research collaborations are built on mutual respect, cooperation, trust, and communication. Nevertheless, a collaboration and data sharing agreement between collaborating research partners may be useful to clarify terms of (co-)ownership and (joint) responsibility for research data. Throughout years, our research group has maintained such an agreement to facilitate transparency and fairness for treatment of data, to ensure compliance with legal and ethical obligations and to ensure that parties take appropriate technical and organisational measures to protect the security and confidentiality of data.

This agreement has evolved to also clarify roles and responsibilities for processing personal information under POPIA and GDPR. POPIA stipulates that South African research institutions may only transfer personal information if the 'third party' is subject to a 'law, binding corporate rules or binding agreement' (Section 71 (1)(a)) which provide an adequate level of protection for the handling of the personal information. In the context of our research projects, this agreement outlines the parameters through which data from children, adolescents, and young people can be shared between the collaborating institutions and which legislations govern this. Therefore, this agreement demonstrates that South African research data may be transferred to a third party governed by GDPR, which provides an appropriate level of protection for the personal information of data subjects. It is important to signal a few important elements in this instrument. First, it is set up as a contractual document among all relevant parties, therefore institutional representatives need to be involved. It is important that the agreement clarifies names of entities and differentiates between the research collaboration and individual studies. For example, the agreement may include a 'memorandum of understanding' template which individual studies may use to enter a collaboration with NGOs for research. Additional addenda may be included in reference to processes unique to individual studies governed the agreement. Additionally, a 'data use undertaking' template may be included to ensure that both parties use consistent terms for sharing data with external data users.

Instrument 7: Data collection, processing, and storage

POPIA highlights the importance of implementing secure measures to protect personal information throughout data collection, processing, and storage. To do this effectively, this must be adequately resourced and budgeted for, and additional support may be required from services within research institutions and technical experts.

During restrictions imposed by the COVID-19 pandemic, researchers have transitioned from face-to-face to remote data collection, enhancing the digitalisation of data collection processes. This demands the implementation of proportionate security measures, and our team uses reliable open-sourced data collection platforms such as REDCap and Open Data Kit. Both have sufficient technical capabilities and functionalities for data collection processes with end-to-end encryption technologies. Research institutions may have their own processes for evaluating the level of security of third-party services, devices, and tools, reducing demand on research teams to resource this expertise internally. Both the security and information protection compliance should be assessed before use. Low-tech techniques (e.g. concealing sensitive information by using unique identifiers or pseudonyms) may also be used to ensure security of personal data. Electronic data captured should be submitted to servers daily, and data should be encrypted between the data collection device and the data servers. Finally, specific protocols may be developed to support POPIA compliant governance of data collection, processing, and storage, defining procedures and parameters for: (1) data retention, (2) data de-identification, and (3) management of access credentials to enable responsible parties to establish common and compliant standards. This data management work needs to be well resourced, as it incorporates additional layers of research governance required by POPIA and data protection regulations beyond South Africa.

Conclusion

POPIA presents an opportunity for researchers to further safeguard the rights and interests of research participants. Although POPIA must be read and applied alongside other relevant legislation for research[11], we advocate that researchers consider adjusted instruments to protect the personal information of children and adolescents that are consistently applied to all research projects. The seven instruments outlined here should be taken as complementary and adaptive and are a response to the risks brought by the increased datafication in research. Despite challenges faced by their implementation, including the increased resource needs, we share them as examples of positive practices, to densify the 'Discussions on POPIA' series, and to achieve the wider goal of safeguarding children's and adolescents' personal information in research.

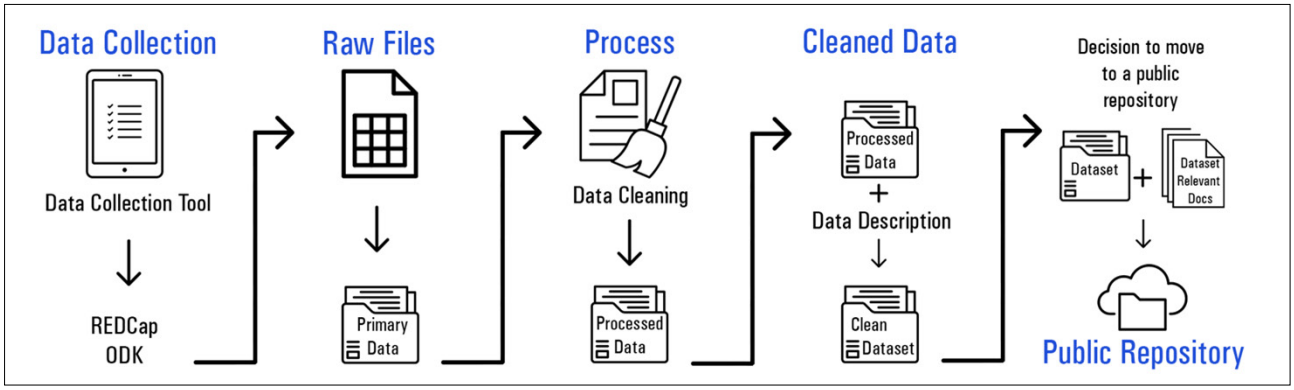


Figure 2: Data workflow from collection to open access repository.

Acknowledgements

We thank all personnel at the University of Cape Town (UCT) and University of Oxford involved in designing, implementing, and improving the instruments outlined here. We also thank Dr Ciara Stanton for sharing her incredible knowledge of POPIA and its applications to research, and Professor Gayle Sherman for spurring our team's thinking around POPIA and our research. We also thank UCT's Library Services for providing rich information that fed Table 1, and Olusegun Akinwande for allowing us to adapt and use his data workflow presented here in Figure 2. We are particularly grateful to the children, adolescents, and their caregivers, who continuously engage with our research projects and give meaning to our research endeavours.

We thank our funders: (UKRI GCRF) Accelerating Achievement for Africa's Adolescents (Accelerate) Hub (Grant Ref: ES/S008101/1); the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 771468); UNICEF Eastern and Southern Africa Office (UNICEF-ESARO), Oak Foundation [R46194/AA001; OFIL-20-057]; the National Institute of Health (NIH)/ Fogarty International Center, National Institute on Mental Health, under Award Number K43TW011434. The content is solely the responsibility of the authors and does not represent the official views of the National Institutes of Health; the Medical Research Council (MRC) and the Department of Health and Social Care (DHSC) through its National of Health Research (NHR); CIPHER grant from International AIDS Society [2018/625-TOS].

Competing interests

We have no competing interests to declare.

Authors' information

The authors are all members of the Research Governance team at the UKRI GCRF Accelerating Achievement for Africa's Adolescents (Accelerate) Hub, co-led by the Universities of Cape Town and Oxford.

References

1. Sawyer SM, Azzopardi PS, Wickremarathne D, Patton GC. The age of adolescence. Lancet Child Adolesc Health. 2018;2(3):223–228. http://dx.doi.org/10.1016/S2352-4642(18)30022-1
2. Statistics SA. Mid-year population estimates. Stat Release. 2020;P0302.
3. Adams R, Adeleke F, Anderson D, Bawa A, Branson N, Christoffels A, et al. POPIA Code of Conduct for Research. S Afr J Sci. 2021;117(5/6), Art. #10933. https://doi.org/10.17159/sajs.2021/10933
4. Cluver L, Boyes M, Bustamam A, Casale M, Henderson K, Kuo K, et al. The cost of action: Large scale, longitudinal quantitative research with AIDS-affected children in South Africa. Ethical Quandaries Soc Res. 2015:41–56.
5. Hensen B, Mackworth-Young CRS, Simwinga M, Abdelmagid N, Banda J, Mavodza C, et al. Remote data collection for public health research in a COVID-19 era: Ethical implications, challenges and opportunities. Health Policy Plan. 2021;36(3):360–368. https://doi.org/10.1093/heapol/czaa158
6. Bostock L. "God, she's gonna report me": The ethics of child protection in poverty research. Child Soc. 2002;16(4):273–283. https://doi.org/10.1002/chi.712
7. Strode AE, Slack C. Selected ethical-legal norms in child and adolescent HIV prevention research: Consent, confidentiality and mandatory reporting [revised]. European and Developing Countries Clinical Trials Partnership (EDCTP); 2012.
8. Strode AE, Slack C. Child privacy rights: A 'Cinderella' issue in HIV-prevention research. South Afr J HIV Med. 2013;14(3):108–110. https://doi.org/10.4102/sajhivmed.v14i3.62
9. Khuluvhe M. Adult illiteracy in South Africa. Pretoria: South African Department of Higher Education and Training; 2021.
10. He E, Hertzog L, Toska E, Carty C, Cluver L. mHealth entry points for HIV prevention and care among adolescents and young people in South Africa. CSSR Working Paper [document on the Internet]. c2020 [cited 2020 Sep 21]. Available from: http://rgdoi.net/10.13140/RG.2.2.36698.98241
11. Swales L. The Protection of Personal Information Act and data de-identification. S Afr J Sci. 2021;117(7/8), Art. #10808. https://doi.org/10.17159/sajs.2021/10808

AUTHOR:
Roseanne Diab[1,2]

AFFILIATIONS:
[1]Director: GenderInSITE, Trieste, Italy
[2]Emeritus Professor, Environmental Sciences, University of KwaZulu-Natal, Durban, South Africa

CORRESPONDENCE TO:
Roseanne Diab

EMAIL:
diab@ukzn.ac.za

HOW TO CITE:
Diab R. Taking stock of climate change science and technology in South Africa: Insights, recommendations and missed opportunities. S Afr J Sci. 2021;117(9/10), Art. #12292. https://doi.org/10.17159/sajs.2021/12292

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
climate change mitigation and adaptation, capacity development, funding instruments, concentrated solar power, South Africa

PUBLISHED:
29 September 2021

Taking stock of climate change science and technology in South Africa: Insights, recommendations and missed opportunities

The formal publication of the *Second Biennial Report on the State of Climate Science and Technology in South Africa*[1] by the Academy of Science of South Africa (ASSAf) in early August 2021 comes ahead of the Conference of Parties (COP) meeting to be held in Glasgow in November 2021. The 26th COP meeting is where world leaders, scientists and representatives of business and civil society will gather to find bold solutions and limit the risks of climate change.

Faced with these imperatives, an assessment of the state of South Africa's research and development activities relating to climate change adaptation and mitigation is timely and vital to understanding our country's readiness to respond to this global challenge.

The second biennial report builds on the first report that was completed at the end of 2016 and covered the period 2005–2015. The second report utilises data from the period 2016 to 2017. Although the report was finalised in 2019, it has taken almost 2 years for government to approve its release into the public domain.

The report presents statistics on key indicators such as the production of master's and PhD graduates, publication outputs and investment into climate-related science and technology (S&T). Such metrics allow for the determination and analysis of trends – one of the most critical contributions of a status report. If they are to be useful, status reports must be produced regularly on an on-going basis. A 2-year interval may be too demanding, given some of the data collection complexities; in the future, a 5-year interval might be more manageable and affordable.

Over and above the presentation of data on key indicators mentioned above, in this second biennial report there was a specific focus on sources and levels of investment into climate change S&T. ASSAf commissioned two studies to survey sources and levels of investment accessed through climate change international finance instruments and from the private sector. This information was complemented with information from annual reports of companies and state-owned enterprises, as well as key informant interviews, to compile a more comprehensive funding source evaluation. As a result, there was an upward revision of investment into climate-related S&T between the first and second biennial reports, with estimates increasing from ZAR400 million per annum (p.a.) to ZAR900 million p.a. Additional funding came predominantly from 'green funds' mobilised by the private sector and support from international financial instruments. It is possible that ZAR900 million p.a is still an underestimate and the ASSAf report calls for more systematic collection of information on investment and finance by the private sector.

The decision to combine reporting on a standard set of indicators with an in-depth analysis of one aspect – in this case, finance – is a good model that should be replicated in the future. For example, there is a need for a more nuanced and in-depth analysis of human capacity development. While it is good to know the numbers of master's and PhD graduates in climate change that South Africa is producing, it would be useful to know whether there are subject area gaps that need to be addressed, what the graduates' employment uptake is, whether their skills set matches the needs of the market, and how many are engaged actively in research and development.

A striking, rather sensitive, finding that is not given prominence in the Executive Summary is the dominance of universities in accounting for climate change related research outputs compared with the non-university sector. Universities account for 85% of research outputs. Government-funded research councils trail far behind, with the highest ranked, the Council for Scientific and Industrial Research, only coming in at 10th position in terms of ranking and producing only a quarter of the output of the highest ranked university. This is cause for concern given the additional teaching responsibilities of university staff and signals an area where improvements can be made.

One of the greatest challenges of producing a status report is demarcating the boundaries of climate change S&T. Unfortunately, there is no right or wrong answer here. It cannot be that the term 'climate change' must appear in the project title or even in the abstract, for climate change has such a far-reaching impact. Climate change can act as justification for research, innovation and investment into renewable energy, and at the same time be relevant for research into drought-resistant crops. What is important, however, is clarity on the approach adopted, specification of the keywords used and consistency in reporting over time. This is certainly the case here. The report utilised hundreds of keywords based on international best practice and tended towards a broad, inclusive approach that is probably advisable given the cross-cutting nature of climate change.

The in-depth analysis of finance instruments was revealing. Of the 37 international instruments identified to fund investments for climate change mitigation and adaptation, South Africa managed to access funding from only 16. There is a missed opportunity here that led to the recommendation that government should publish a register of international funding instruments and disseminate it to research institutions. It was also revealing that South Africa dominated the funding pattern for the entire sub-Saharan Africa, accessing more than one-third (36.9%) of funds approved. However, rather than congratulating ourselves for dominating over other less-advantaged African countries, the overriding message should be the underutilisation of global funding opportunities.

Two interesting recommendations are made, both of which relate to South Africa's strategic positioning in low-carbon energy technologies. The first is the identification of concentrated solar power as a technology that is ideally suited to South African conditions and skills. South Africa could become a global player according to

the report. A similar recommendation was made in a previous report on *The State of Energy Research in South Africa*[2]. It is estimated that 60% of concentrated solar power systems could be manufactured locally. A second opportunity relates to the local manufacture of fuel cells, particularly where they require platinum-group metals, given South Africa's vast reserves. These are clear attempts to provide direction to future government investments that will leverage local advantage.

This second biennial report on climate change S&T is a useful addition to the South African climate change literature. It allows one to take stock and provides an evidence base for improvements. It is sincerely hoped that the South African Department of Science and Innovation will continue this tradition, but more importantly, that the recommendations made will be carefully considered and implemented where possible.

Competing interests

I have no competing interests to declare. I was a member of the ASSAf panel responsible for the *Second Biennial Report on the State of Climate Science and Technology in South Africa*.

References

1. Academy of Science of South Africa (ASSAf). Second biennial report on the state of climate science and technology in South Africa. Pretoria: ASSAf; 2019. http://dx.doi.org/10.17159/assaf.2019/0038
2. Academy of Science of South Africa (ASSAf). The state of energy research in South Africa. Pretoria: ASSAf; 2014. http://dx.doi.org/10.17159/assaf/0026

AUTHORS:
Sam M. Ferreira[1,2]
Luthando Dziba[3]

AFFILIATIONS:
[1]Scientific Services, SANParks, Skukuza, South Africa
[2]Department of Conservation and Marine Sciences, Cape Peninsula University of Technology, Cape Town, South Africa
[3]Conservation Services Division, SANParks, Pretoria, South Africa

CORRESPONDENCE TO:
Sam Ferreira

EMAIL:
sam.ferreira@sanparks.org

HOW TO CITE:
Ferreira SM, Dziba L. Where are rhinos safest? S Afr J Sci. 2021;117(9/10), Art. #11300. https://doi.org/10.17159/sajs.2021/11300

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
rhinoceros, private owners, provincial reserves, community owners, national parks

PUBLISHED:
29 September 2021

Where are rhinos safest?

During 2020, South Africa reported 394 poached rhinoceros (rhino) carcasses – 34 at private properties and 360 at state properties.[1] Poaching since 2008 has degraded South Africa's rhino populations. By the start of 2020, South Africa had 14 410 southern white rhinos (*Ceratotherium simum simum*) and 1900 black rhinos, mostly south-central (*Diceros bicornis bicornis*) and south-western black rhino (*D. b. minor*) sub-species and a few non-native eastern black rhinos (*D. b. michaeli*) (Table 1).

Rhinos occur in national parks managed by SANParks, reserves managed by provincial conservation authorities and non-state areas comprising private and communal ownership. Of these, the Kruger National Park (Kruger) has experienced the highest number of poaching losses reported for state land. Kruger has an operational area the size of Israel (approximately 2 000 000 ha). In contrast, privately owned nature reserves containing rhinos in South Africa have a national average of 10 000 ha (1 300 000 ha distributed across 132 properties studied[2]). Therefore, privately owned reserves are, on average, approximately 200 times smaller than the size of Kruger. National parks other than Kruger, provincial and communal reserves are also smaller relative to Kruger, with the other national parks averaging approximately 33 000 ha, excluding Addo Elephant National Park (Addo), which is 173 000 ha. Addo, however, has five disconnected sections, three of which have rhinos present.

The recently released report by a Ministerial High-level Panel on the management and use of lion (*Panthera leo*), leopard (*P. pardus*), elephants (*Loxodonta africana*) and rhinos made several explicit recommendations.[3] Key aspects reflect on a preference for disinvesting in captive rhino breeding operations and investing in secured, restored, and rewilded natural landscapes containing, amongst others, thriving rhino populations, despite a considerable diversity of concepts of wild and rewilding.[4]

What lessons can South Africa derive from different successes across the different scales, management approaches and land uses associated with rhinos? Size of the area may be a key factor that influences the ease of poachers finding rhinos to poach and rangers finding rhinos to protect. A meaningful evaluation of safety for rhinos may be to compare national parks (other than Kruger) with both provincial reserves and private reserves. The average size of reserves in these three categories is comparable, thus allowing direct assessment of consequences of anti-poaching activities and rhino conservation.

Enriching our understanding of rhino protection could further benefit from scaling carcasses by the number of rhinos living within these different categories of land uses and management authorities (Table 1). All three categories – national parks other than Kruger, provincial reserves and private reserves – do much better than does Kruger itself. When comparisons account for rhino population sizes, poaching rates in Kruger were 6.1%, the lowest since 2013[5], but 2.4 times as high as those in provincial reserves (2.6%), 13.6 times higher than those for non-state reserves (0.5%) and 20.4 times higher than in national parks excluding Kruger (0.3%) during 2020. These results directly contrast the statement from the Private Rhino Owners Association leadership recently reported in the media[6]:

> *In South Africa, if you are a rhino on a state reserve, your chances of getting whacked by a poacher are about nine times greater than if you roam a private one.*

The size of focal areas does play an important role. For instance, the number of rhinos poached per population per unit area in Kruger is higher than that in other national parks (4.1 times) and non-state reserves (7.8 times).

Unintended messaging may have had a boomerang effect on rhino conservation initiatives[7], such as exposing critical security intelligence utilised by poaching syndicates. For instance, media framed the translocation of rhinos to strongholds outside Kruger as a 'rhino evacuation'[8]. Poaching soared in the weeks that followed as traffickers sought to cash in before the opportunity to do so diminished. Conservation agencies always battle with the democratic rights of citizens to information and many authorities have limited experience in managing sensitive intelligence. Similarly, a recent media report in 2021 implied broadscale incompetency in state conservators of rhinos in South Africa.[2] Such misleading messaging is an even greater challenge, and flaunts basic standard practices for media reporting such as providing balanced perspectives and giving affected parties the right to comment.[9] For instance, the report on state incompetency in rhino conservation noted one source of information, and the reporter did not provide responses from alternative sources to get other perspectives that would have allowed richer engagement with the information available in the public domain.

Our brief analyses highlight that responsible and balanced media reporting could facilitate lessons stakeholders can collectively learn by looking at, for instance, what makes non-state reserves, provincial reserves and small national parks more successful compared to Kruger in managing the threat of poaching to rhino populations. The low poaching rates may reflect key aspects of small national parks and non-state reserves – the latter typically well resourced, usually at own cost by private industry[3], but importantly may have optimal operational areas that allow operational efficiency in situational awareness, access control, staff integrity management and intensive monitoring of the rhino assets themselves in addition to standard anti-poaching operations. Small provincial reserves may have optimal sizes, but have limited resources and/or have challenges in operational efficiency that may compromise situational awareness, access control, integrity and knowledge of the rhino asset. Kruger is well resourced, but is extremely large beyond the optimal size to gain effective anti-poaching control. The large operational area most likely constrains Kruger's ability to achieve operational efficiency in situational awareness, access control, integrity management and individual level knowledge of the rhino assets despite impressive anti-poaching operations.

Combining local insight into successes in rhino safety with insights gained from similar conservation areas with rhinos elsewhere in Africa should enhance interventions that result in low poaching rates.[10] These continental insights highlighted features like leadership and integrity management as well as intensive monitoring of rhino assets. In-depth reflection on comparable poaching rates suggest that South Africa could benefit from approaches that innovatively distribute resources appropriately through various means to those places that are of optimal sizes with all associated operational efficiency benefits. Identification of mechanisms to generate funds for rhino protection initiatives typically suffers from restrictive arguments for international trade in horn[11] when few protected areas and agencies make use of the full suite of conservation-funding models and incentives[12].

Another key insight is that authorities may improve rhino safety by 're-sizing' Kruger into rhino zones of reduced size, closer to the optimality of private, provincial and other state parks. At present, Kruger has administrative sections that mismatch rhino distribution. The definition of different zones based on environmental characteristics, such as rhino safe havens and ecological traps[13], may thus re-ignite the benefits of the Intensive Protection Zone, Joint Protection Zone and Composite Protection Zone initiated when Kruger had 10 000 rhinos[14].

Global analyses predict that governance capacity may actually dictate biodiversity conservation success[15], similar to recent media reports[2]. Using this approach in our analysis shows that provinces with low poaching rates (excluding Kruger) actually showed variable governance success as indicated by the 2018–2019 audit status of provincial level municipalities[16] (Table 1). Provinces with substantially higher poaching rates, such as the North West, Limpopo and Mpumalanga, had generally lower governance success. KwaZulu-Natal provided the greatest contrast with patterns in governance challenges – despite relatively high indicators of governance success, the poaching rates were the highest in the country excluding those of Kruger. A more useful outcome would be an analysis of which factors allow the achievement of low poaching rates despite governance challenges in certain provinces, to allow their replication in other areas.

Although South Africa embraces human democratic rights-based conservation ideologies[17], explicit mandates introduce different risks to private industry, a key element of non-state ownership of rhinos, and state-managed reserves and parks. Private industry does not carry the burden of a constitutional mandate that imposes an explicit requirement to protect South Africa's heritage, irrespective of state resources available. The private sector possesses impressive business acumen, but business outcomes of their rhino conservation efforts depend on the broader business environment. The onslaught of poaching imposes significant costs on private industry and has been recognised as one of the main drivers of disinvestment in rhinos – by 2018, 21% of private rhino owners were disinvesting[3]. Between 2012 and 2017, ten state reserves lost or removed white rhino and an additional three lost or removed black rhino.[18] Removals to alternative safer state reserves, however, did not divorce state agencies from their constitutional mandate of rhino conservation. For many detractors, these trends of disinvestment could fuel a claim that 'A rhino living in a private reserve has a higher chance of a forceful eviction from home than one on a state reserve'. Yet, translocation of animals to safer areas is a rational response in the face of the poaching onslaught driven by the criminal activity of horn trafficking. Dealing with trafficking remains a national and international law enforcement challenge.

Unintended messaging may further serve to demotivate state and private reserve managers working hard under varying constraints. It is an extraordinary achievement of private industry to be playing such a major role in the conservation of rhinos and in protecting large fractions of South Africa's rhinos (Table 1; black rhino – 23.7%, white rhino – 48.6%).

Table 1: Population sizes at the beginning of 2020 and poaching rates within four categories of land uses associated with rhino protection in South Africa during 2020. Rates within provinces contribute relatively differently to countrywide poaching rates associated with the four categories of land uses related to rhino protection. When we focus only on poaching in a province, incidences do not associate with an index of governance quality.[15]

	Kruger National Park (KNP)	Other national parks	Provincial reserves	Private properties	Rhino properties excluding KNP	Municipal governance index[a]
Number of black rhinos[b]	270	315	865	450	1630	
Number of white rhinos[c]	3550	360	3500	7000	10 860	
Poaching rates	6.12%	0.30%	3.10%	0.50%	1.30%	
Free State			0.000%	0.000%	0.000%	56.0%
Northern Cape		0.000%	0.000%	0.015%	0.009%	58.6%
Eastern Cape		0.000%	0.000%	0.000%	0.000%	67.7%
Gauteng			0.000%	0.029%	0.018%	82.2%
Western Cape		0.000%	0.000%	0.015%	0.009%	87.6%
North West			0.466%	0.029%	0.166%	38.8%
Mpumalanga	6.073%		0.110%	0.132%	0.114%	58.9%
Limpopo	0.047%	0.300%	0.027%	0.250%	0.175%	65.0%
KwaZulu-Natal			2.496%	0.029%	0.814%	71.3%

[a]Provincial indices based on a weighted average derived from the complete audit qualification of municipalities during 2018–2019. We assigned each municipality with values reflecting the audit outcome as follows: 0 – Outstanding audits, 1 – Disclaimed with findings, 2 – Adverse with findings, 3 – Qualified with findings, 4 – Unqualified with findings, 5 – Unqualified with no findings. By averaging these scores per province excluding municipalities with outstanding audits, and expressing the result as a percentage of 5, we obtained an index of municipal governance quality.

[b]Summary of black rhino data extracted from the SADC Rhino Management Group that provided estimates for 2015 and 2017 separated into national parks, provincial and private reserves. See Emslie et al.[18] We used the trends in provincial and private reserves between 2015 and 2017 to predict likely numbers at the beginning of 2020. For national parks, the estimates were based on formal survey records at the end of 2019.

[c]Summary of white rhino data provided by the Scientific Authority.

Indeed, South Africans should celebrate and acknowledge the outstanding work that both private sector and managers of some state properties are doing under trying circumstances and constraints. Society, however, should support all reasonable conservation efforts and dissuade those that distract South Africans from complex conservation challenges. An open environment that does not restrict learning and that supports developing new insights into how best to respond to these challenges can change the plight of rhinos and the future of South Africa's biological heritage. We do hope that those who want to tell the latest rhino stories would also do their best to report them in a responsible manner. The lesson is not about who keeps rhinos safest. It is about what is the safest way to keep rhinos. Even so, multiple approaches associated with rhinos can provide beneficial outcomes to support South Africa's rhino conservation initiatives and reputation.

Acknowledgements

This Commentary benefited from discussions with Jo Shaw, Johan Jooste, Danie Pienaar, Danny Govender and Stef Freitag-Ronaldson.

Competing interests

We have no competing interests to declare.

References

1. South African Department of Forestry, Fisheries and the Environment (DFFE). Environment, Forestry and Fisheries report back on rhino poaching in South Africa in 2020 [webpage on the Internet]. c2021 [cited 2021 May 21]. Available from: https://www.gov.za/speeches/environment-forestry-and-fisheries-rhino-poaching-south-africa-2020-1-feb-2021-0000#
2. Clements HS, Knight M, Jones P, Balfour D. Private rhino conservation: Diverse strategies adopted in response to the poaching crisis. Conserv Lett. 2020;13(6), e12741. https://doi.org/10.1111/conl.12741
3. South African Department of Forestry, Fisheries and the Environment (DFFE). The High-level Panel of Experts for the review of policies, legislation and practices on matters of elephant, lion, leopard and rhinoceros management, breeding, hunting, trade and handling: High-level Panel report [document on the Internet]. c2020 [cited 2021 May 21]. Available from: https://www.parliament.gov.za/storage/app/media/Pages/2017/october/High_Level_Panel/HLP_Report/HLP_report.pdf
4. Bühne HST. Pettorelli N, Hoffmann M. The policy consequences of defining rewilding. Ambio. 2021. https://doi.org/10.1007/s13280-021-01560-8
5. Ferreira SM, Greaver C, Knight GA, Knight MH, Smit IP, Pienaar D. Disruption of rhino demography by poachers may lead to population declines in Kruger National Park, South Africa. PLoS ONE. 2015;10(6), e0127783. https://doi.org/10.1371/journal.pone.0127783
6. Stoddard E. In the war on rhino poaching, data shows private parks are doing a better job. Daily Maverick. 2021 February 09. Available from: https://www.dailymaverick.co.za/article/2021-02-09-statistics-show-rhino-poaching-in-south-africa-plunged-33-in-2020/
7. Glenn I, Ferreira SM, Pienaar D. Communication on rhino poaching: Precautionary lessons about backfires and boomerangs. S Afr J Sci. 2019;115(3–4), Art. #5585. http://dx.doi.org/10.17159/sajs.2019/5585
8. British Broadcasting Corporation. S Africa rhinos to be evacuated from Kruger National Park. BBC News. 2014 August 12; Africa. Available from: https://www.bbc.com/news/world-africa-28758910
9. Jennings M. The evolution and devolution of journalistic ethics. Imprimis. 1999;28:1–4. Available from: https://imprimis.hillsdale.edu/wp-content/uploads/2016/10/Imprimis-The-Evolution%E2%80%93-and-Devolution%E2%80%93-of-Journalistic-Ethics-Jul-1999.pdf
10. Ball M, Wenham C, Clegg B, Clegg S. What does it take to curtail rhino poaching? Lessons learned from twenty years of experience at Malilangwe Wildlife Reserve, Zimbabwe. Pachyderm. 2019;60:96–104. Available from: https://pachydermjournal.org/index.php/pachyderm/article/view/39
11. Eikelboom JA, Nuijten RJ, Wang YX, Schroder B, Heitkönig IM, Mooij WM, et al. Will legal international rhino horn trade save wild rhino populations? Glob Ecol Conserv. 2020;23, e01145. https://doi.org/10.1016/j.gecco.2020.e01145
12. Lindsey P, Allan J, Brehony P, Dickman A, Robson A, Begg C, et al. Conserving Africa's wildlife and wildlands through the COVID-19 crisis and beyond. Nat Ecol Evol. 2020;4(10):1300–1310. https://doi.org/10.1038/s41559-020-1275-6
13. Le Roex N, Dreyer C, Ferreira SM. Poaching creates ecological traps within an iconic protected area. Anim Conserv. 2020;23(3):250–259. https://doi.org/10.1111/acv.12532
14. Shaw M, Rademeyer J. A flawed war: Rethinking 'green militarisation' in the Kruger National Park. Politikon. 2016;43(2):173–192. https://doi.org/10.1080/02589346.2016.1201379
15. Smith RJ, Muir RD, Walpole MJ, Balmford A, Leader-Williams N. Governance and the loss of biodiversity. Nature. 2003;426(6962):67–70. https://doi.org/10.1038/nature02025
16. Auditor General, South Africa. Auditor-general releases municipal audit results under the theme – "not much to go around, yet not the right hands at the till". Pretoria: Auditor General, South Africa; 2020. Available from: https://www.agsa.co.za/Portals/0/Reports/MFMA/201819/Media%20Release/2020%20MFMA%20Media%20Release%20Final.pdf
17. Madzwamuse M, Rihoy E, Louis M. Contested conservation: Implications for rights, democratization, and citizenship in southern Africa. Development. 2020;63(1):67–73. https://doi.org/10.1057/s41301-020-00237-1
18. Emslie RH, Milliken T, Talukdar B, Burgess G, Adcock K, Balfour D, et al. African and Asian rhinoceroses – Status, conservation and trade. A report from the IUCN Species Survival Commission (IUCN SSC) African and Asian Rhino Specialist Groups and TRAFFIC to the CITES Secretariat pursuant to Resolution Conf. 9.14 (Rev. CoP17). COP18, Doc 83.1, Annex 2 [document on the Internet]. c2018 [cited 2021 May 21]. Available from: https://www.traffic.org/site/assets/files/12220/african-asian-rhinos-iucn-traffic.pdf

AUTHORS:
Graham J. Alexander[1]
Krystal A. Tolley[1,2]
Bryan Maritz[3]
Andrew McKechnie[4,5]
Paul Manger[6]
Robert L. Thomson[7]
Carsten Schradin[1,8]
Andrea Fuller[9,10]
Leith Meyer[9,10]
Robyn S. Hetem[1]
Michael Cherry[11]
Werner Conradie[12,13]
Aaron M. Bauer[14]
David Maphisa[2,15]
Justin O'Riain[16]
Daniel M. Parker[17,18]
Musa C. Mlambo[19,20]
Gary Bronner[16]
Kim Madikiza[1]
Adriaan Engelbrecht[3]
Alan T.K. Lee[7,21]
Bettine Jansen van Vuuren[22]
Tshifhiwa G. Mandiwana-Neudani[23]
Darren Pietersen[5]
Jan A. Venter[13]
Michael J. Somers[24]
Rob Slotow[25]
W. Maartin Strauss[26]
Marc S. Humphries[27]
Peter G. Ryan[7]
Graham I.H. Kerley[28]

AFFILIATIONS:
[1]School of Animal, Plant and Environmental Sciences, University of the Witwatersrand, Johannesburg, South Africa
[2]South African National Biodiversity Institute, Kirstenbosch Research Centre, Cape Town, South Africa
[3]Department of Biodiversity and Conservation Biology, University of the Western Cape, Cape Town, South Africa
[4]South African Research Chair in Conservation Physiology, South African National Biodiversity Institute, Pretoria, South Africa
[5]Department of Zoology and Entomology, University of Pretoria, Pretoria, South Africa
[6]School of Anatomical Sciences, University of the Witwatersrand, Johannesburg, South Africa
[7]FitzPatrick Institute of African Ornithology, University of Cape Town, Cape Town, South Africa
[8]IPHC, CNRS, University of Strasbourg, Strasbourg, France
[9]Brain Function Research Group, School of Physiology, University of the Witwatersrand, Johannesburg, South Africa
[10]Department of Paraclinical Sciences and Centre for Veterinary Wildlife Studies, Faculty of Veterinary Science, University of Pretoria, Pretoria, South Africa
[11]Department of Botany and Zoology, Stellenbosch University, Stellenbosch, South Africa
[12]Port Elizabeth Museum (Bayworld), Port Elizabeth, South Africa
[13]Department of Nature Conservation Management, Natural Resource Science and Management Cluster, Nelson Mandela University, George, South Africa
[14]Department of Biology and Center for Biodiversity and Ecosystem Stewardship, Villanova University, Villanova, Pennsylvania, USA

Excessive red tape is strangling biodiversity research in South Africa

Preventing the over-exploitation of natural resources is vital to ensure that biodiversity is protected and conserved.[1,2] Legislation and regulations are therefore necessary to manage resource utilisation, but overly stringent legislation and regulations can have unintended negative consequences. For example, biodiversity research, much of which is state funded, is now subject to excessive red tape to the extent that overregulation is impeding progress. Researchers must navigate a myriad of laws, rules, permit requirements, ethics clearances and approvals, many of which require annual renewal, progress reporting, and submission of amendment applications for ongoing projects. Excessive red tape particularly hinders field-based research, and in our experience, has a negative impact on research productivity in South Africa. If current levels of bureaucracy and managerialism persist, we believe that the impact on biodiversity research in the country will be debilitating. Former South African Minister of Finance, Tito Mboweni, has acknowledged the negative impact of red tape on small business enterprise and economic growth in South Africa, and there are now attempts to reduce it.[3] So too, excessive red tape and overregulation of research should be rationalised to enhance knowledge generation and application.

Over the last decade, new legislation, new interpretations of established legislation and increasing administrative oversight have massively increased the administrative and compliance burden experienced by researchers in South Africa.[4] This particularly impacts field-based research, which typically requires various permits, permissions, and authorisations for (1) the collection of biological samples on public or private land, (2) performing certain procedures on animals, (3) transporting of biological materials and samples, especially when the transport is across provincial or international borders, as well as (4) the storage and usage of samples.

Most field-based research projects have a strong conservation theme, and their findings inform conservation policies and management so that South Africa's biodiversity can be protected effectively (e.g. as required by the *National Environmental Management: Biodiversity Act* (NEMBA) *No. 10 of 2004*).[5] Ironically, the legislation, managerialism and bureaucracy which are purported to be enacted for the very purpose of protecting South Africa's remarkable biodiversity are collectively now having a negative impact on conservation by hindering research. Overregulation of foundational biodiversity research has other knock-on effects which are detrimental to the achievement of national strategic goals. For example, undergraduate research training and skills development are enhanced through practical work in the field, which is negatively impacted because overwhelmed academics avoid field-based teaching due to the onerous regulatory framework that must be navigated and the unpredictable delays in permit approval. The number of postgraduate students managed and supervised by academic researchers is also curtailed by the administrative and compliance burden, slowing the development of local capacity and transformation, and extending the time taken for students to complete their postgraduate degrees. Slowed student throughput rates impact government subsidies to universities, further retarding capacity building and making biodiversity research less attractive as a career path.[2]

Red tape comes in many forms, and researchers must deal with it at many levels. In South Africa, more than a dozen different National Acts and accompanying regulations, which are regularly revised, can directly impact any field-based biological research project[4], requiring permits that may take months to be issued. Additionally, provincial regulations also require researchers to apply for permits for several activities, and so research programmes may require several permits from any given province. Moreover, broadscale projects conducted over more than one province require permits from each of the relevant provinces, each with its own permitting system and set of rules[6], with some requiring other permits to be in place before an application is considered. Thus, it is not uncommon for some field-based research projects to require upwards of 20 different permits, clearances and approvals to be issued before work can commence.[4]

Another layer of regulation comes with the requirements for animal ethics clearance. Ethics committees are constituted in accordance with directives outlined in the South African National Standards document (SANS 10386) and are generally administered by universities or research institutes that employ researchers. As stipulated in SANS 10386, committees are composed of veterinarians, animal researchers, representatives of welfare organisations and lay persons, and are now audited and accredited by the National Health Research Ethics Council (NHREC). However, some organisations that manage land where research is carried out (e.g. South African National Parks) have constituted their own animal use committees and do not accept clearance certificates from other NHREC-accredited committees. Thus, some collaborative research projects may require clearance from several ethics committees even when each is accredited by the NHREC. This has been further compounded by the National Research Foundation which now requires ethics clearance to be in place prior to the release of funds. In the case of student funding, ethical clearance must be in the student's name, leading to further duplication of clearances needed. Because students must register for their degree prior to applying for ethics clearance (which usually takes several months), they may be stranded without funding for an extended time or may even lose their bursaries if they miss deadlines.

In our experience, unjustified delays in the issue of one or two permits may hold up a research project to the extent that other permits, which are usually valid for one year, lapse before work can even begin. Such delays jeopardise research funding because many funding bodies maintain tight funding regimes and monies must be returned if not spent within a funding cycle. Over the last decade, administrative oversight at the various levels of legislation has ballooned to the extent that researchers now spend a significant portion of their research time on legislative compliance. These complicated procedures, inefficiencies, and delays in the issuing of permits often foil research progress.

[15]Statistics in Ecology, Environment and Conservation, Department of Statistical Sciences, University of Cape Town, Cape Town, South Africa
[16]Institute for Communities and Wildlife in Africa, Department of Biological Sciences, University of Cape Town, Cape Town, South Africa
[17]School of Biology and Environmental Sciences, University of Mpumalanga, Mbombela, South Africa
[18]Wildlife and Reserve Management Research Group, Department of Zoology and Entomology, Rhodes University, Makhanda, South Africa
[19]Department of Freshwater Invertebrates, Albany Museum, Makhanda, South Africa
[20]Department of Zoology and Entomology, Rhodes University, Makhanda, South Africa
[21]Centre for Functional Biodiversity, University of KwaZulu-Natal, Pietermaritzburg, South Africa
[22]Centre for Ecological Genomics and Wildlife Conservation, University of Johannesburg, Johannesburg, South Africa
[23]Department of Biodiversity, University of Limpopo, Polokwane, South Africa
[24]Mammal Research Institute, Centre for Invasion Biology, University of Pretoria, Pretoria, South Africa
[25]School of Life Sciences, University of Kwazulu-Natal, Pietermaritzburg, South Africa
[26]Department of Environmental Sciences, University of South Africa, Johannesburg, South Africa
[27]School of Chemistry, University of the Witwatersrand, Johannesburg, South Africa
[28]Centre for African Conservation Ecology, Nelson Mandela University, Gqeberha, South Africa

CORRESPONDENCE TO:
Graham Alexander

EMAIL:
graham.alexander@wits.ac.za

HOW TO CITE:
Alexander GJ, Tolley KA, Maritz B, McKechnie A, Manger P, Thomson RL, et al. Excessive red tape is strangling biodiversity research in South Africa. S Afr J Sci. 2021;117(9/10), Art. #10787. https://doi.org/10.17159/sajs.2021/10787

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
research productivity, human capital development, Nagoya Protocol, bureaucracy

PUBLISHED:
29 September 2021

As field researchers, we acknowledge the need for regulations relating to the use of South Africa's natural resources for research and other purposes: legislation is necessary to prevent unethical practices, ensure animal welfare, halt the unsustainable harvest of natural resources, check the spread of notifiable diseases, and curb the illicit wildlife trade. But the implementation of the legislation in terms of legitimate research has become problematic because it is applied with a broad brushstroke approach. In essence, hunters, wildlife poachers and bona fide researchers are viewed through the same legislative lens. This inclusive approach supposedly reduces risks to natural resources, but also stops or retards many genuine research activities that are intended to benefit conservation. Moreover, a broad brushstroke approach is not always effective: to circumvent regulations, the illegal wildlife trade has simply moved underground, while bona fide researchers suffer the consequences of these restrictions. Researchers are easily audited because their work is published in the scientific literature, and permit and certificate numbers must be declared as standard practice. The result is that research is impeded, while illegal wildlife traders evade the regulations.

The explosion of red tape hindering research is not limited to South Africa[7,8] and new global agreements such as the Nagoya Protocol regulate commercial research and development internationally. In the case of the Nagoya Protocol, exchange of biological samples is prohibited unless an Access and Benefit Sharing Agreement exists between participating countries.[9,10] Although the stated intention is admirable and aimed at providing indigenous biological resources with much-needed protection from commercial exploitation, the broad brushstroke approach means that bona fide research material, which is not intended for commercialisation, is included in the restrictions, greatly hindering international research collaborations. Commercialisation of biodiscovery has, in any event, been shown to be limited and usually involves widespread organisms which occur across several countries.[8] More than anything else, the Nagoya Protocol is likely to stifle both research and the sustainable commercial use of natural resources through red tape inertia.

Parties to the Convention on Biological Diversity, the umbrella agreement for the Nagoya Protocol, are currently negotiating approaches to enable benefit sharing from the use of the collection of "digital sequence information on genetic resources" (i.e. DNA sequences). Some of the options currently being considered would result in restrictions on the use of digital sequence information, even for basic non-commercial research unless an Access and Benefit Sharing Agreement was in place.[11,12] This would make phylogenetic analyses of taxa that occur across several countries practically impossible. Such phylogenetic studies form the basis of modern-day taxonomy, systematics and the assessment of biodiversity, which in turn provide the foundational data on which conservation biology rests. Without the necessary field work, tissue sampling and sequencing, cryptic species go undetected and the effectiveness of conservation is reduced.

There are also clear examples in South Africa where 'old' legislation has been reinterpreted with disastrous consequences for research. In some cases, researchers in South Africa are now effectively held hostage by bizarre interpretations of legislation. Here, we highlight two examples: Section 20 of the *Animal Diseases Act No. 35 of 1984* (Box 1) and the *Veterinary and Para-Veterinary Professions Act 19, 1982* (Box 2). In our opinion, these Acts are no longer being interpreted or enforced in the spirit with which they were intended, with dire consequences for biological research. We thus appeal to those who oversee, interpret, and implement the laws and legislation in South Africa to moderate their risk averse approach to facilitate and promote biological research. We advocate a return to a more reasonable and fair interpretation of existing legislation so that scientific endeavour is facilitated and promoted, rather than impeded and blocked. Here, we provide recommendations which we believe would facilitate research without reducing the effectiveness of the legislation in the protection of South Africa's natural resources and biodiversity:

- Legislation should be assessed by an independent expert panel with input from researchers and legislators to facilitate rational and fair interpretation that reflects the intended spirit of the legislation. Where appropriate, permitting regulations should include well-defined exemptions for bona fide research conducted by researchers of good standing and affiliated with accredited research institutions.
- Provincial and national permitting bodies should provide blanket research permits to accredited research institutions. Permission for individual research projects should then be devolved to each institution's ethics committee, which rigorously evaluates all research applications (with input from committee members who have a great deal of experience and knowledge in science, animal-based research and veterinary practice). In addition to reducing delays for permitting of research projects, permission issued in this way would also serve to alleviate pressure on provincial permitting authorities, allowing for faster permit application processing by provinces. Threatened or protected species could be excluded from this process, thus allowing provincial authorities to regulate these species more directly.
- Where permits are required for individual research projects, they should be issued for the expected duration of the project – not on an annual basis as is the current norm. Not only will a longer validity reduce administrative burden on researchers and provincial administrators, but it can also be argued that it is unethical and untenable to embark on a research project where there is no guarantee that there will be provincial permission to complete the research.
- Issuing authorities should apply provisions made in NEMBA (Section 92), which states that the relevant authorities should exercise their powers collectively and issue a single integrated permit inclusive of all aspects of the relevant research in a research proposal where appropriate (e.g. including collection, transport, storage, and transfer internationally), instead of multiple separate permits and authorisations. This provision could dramatically streamline the issuing of research permits, but to our knowledge, it has never been applied by permitting authorities.

- Research permits should routinely include provision for the collection of serendipitously discovered biological samples which are important for documenting occurrence of cryptic or rare animals. For example, records of rare reptiles may be discovered as roadkill. Currently, these specimens may not be collected unless specific permits are already in place.
- Permitting procedures should be streamlined. In our experience, turnaround times are far longer than those promised and appear to be due to unwieldy systems and procedures. We are, however, pleased to note that some provinces have addressed this, and in some instances, the permitting process is reasonably efficient and timely.
- Clearance from an NHREC-accredited ethics committee in South Africa should be valid nationally – there should never be a need for multiple ethics clearances for a single research project.
- Universities and national research institutes should support researchers more directly, for example, with the provision of compliance officers familiar with the pertinent legislation to assist with compliance issues. They should also ensure that research and ethics committees are well equipped, functional and provide streamlined procedures to facilitate ethical research.
- The South African Department of Agriculture, Land Reform and Rural Development should compile a list of 'Section 20' exempt taxa, sample types and study types (see Box 1).
- The South African Veterinary Council (SAVC) should streamline the process of authorisation for procedures, identify a list of exempt procedures in consultation with other relevant professional bodies and their constituents, and reassess the requirement of annual renewal of authorisation for researchers who are not registered with SAVC or the Health Professions Council of South Africa (see Box 2). If the intent of the process of authorisation is to ensure that only competent practitioners perform procedures, the requirement for annual renewal makes little sense.

Scientific research is one of the cornerstones of human progress, development and sustainability, and should therefore be promoted and facilitated by legislation. Biodiversity research informs foundational science, conservation and the management of biodiversity.[5] Although legislators and rule-makers may not be trying to overtly restrict research, we believe that a narrow focus on regulations in their area of influence means that the wider implications of the cumulative impact of the excessive burden of all legislation on researchers is not evident to them. Researchers, on the other hand, have to bear the brunt of increasing bureaucracy and managerialism across the board. If research, and consequently conservation of biodiversity, are to be prioritised, we need the red tape cut as a matter of urgency. Furthermore, conversations with academic colleagues in other disciplines suggest that the negative impacts of bureaucracy and managerialism are not limited to biological sciences, or indeed only to science. Similar issues appear to impact several disciplines in the humanities and health sciences.

Box 1

The *Animal Diseases Act (No. 35 of 1984)* aims to control the spread of animal diseases and generally promote animal health. In Section 20 of the Act, it is stated that a permit is required 'to perform any research, investigation or experiment of any kind for any purpose with or on any animal or parasite or pathogen or part thereof in any form'. Thus, this section, with its all-encompassing and vague definition of biological material, is aimed at curtailing only activities conducted as part of research, while these same activities can be carried out as long as they are not for research purposes. A 'Section 20 permit' for research is granted only through a very detailed and exhaustive application process that has recently become significantly more onerous, and every individual research project that involves an animal or derivate now requires its own Section 20 permit. The wording in Section 20, taken in isolation, has recently been interpreted by the South African Department of Agriculture, Land Reform and Rural Development (DALRRD) to mean that all animal research requires a permit, even if the potential for spreading disease is virtually nil (i.e. collection of tissue samples immediately placed in ethanol or other preservatives for DNA sequencing). Because the intent of the Act relates to controlling animal disease, logically, Section 20 should not apply to animals or samples that cannot carry diseases that require control. Currently, the DALRRD interpretation of the Act as laid out in their Guidelines for Application for a Permit under Section 20 of the *Animal Diseases Act 1984* (Version 20/1) is a wholesale broad brushstroke approach for all animal research work. As part of this process, research laboratories are required to submit various types of additional documentation regarding laboratory operating procedures and biocontainment even where these research laboratories do not investigate animals that require disease control. This Act predates the advent of routine collection of small tissue samples for DNA sequencing purposes, but the rationale has been applied even to this type of research. Thus, a hunter transporting an entire carcass is not impacted by any Section 20 restriction even though the potential for spread of disease is much greater than the collection of a tiny tissue sample sterilised in ethanol. The current interpretation of the Act would even require a dog owner to apply for a Section 20 permit to take the dog for a walk if the owner was counting its steps as part of a research project.

Box 2

The *Veterinary and Para-Veterinary Professions Act, 1982 (Act No. 19 of 1982)* states that only persons registered with the South African Veterinary Council (SAVC) or the Health Professions Council of South Africa (HPCSA) may perform certain procedures on an animal. SAVC has been slow in defining the list of restricted procedures but has recently tabled an exhaustive list of procedures which may only be performed by registered persons (SAVC or HPCSA). Unregistered researchers needing to perform a 'procedure' as part of their research must apply to SAVC for authorisation. This process must be repeated annually, is administratively cumbersome, slow and costly. It has resulted in the need for veterinarians to be available onsite (at significant expense to researchers) to perform even the simplest of procedures (such as injecting an animal or inserting a passive integrated transponder tag) unless the researcher has jumped through the hoops to have their competency evaluated and be authorised by the SAVC to perform the procedure. The requirement for authorisation also means that many procedures can no longer be taught to students in a field setting, impacting the quality and competences of the next generation of researchers. The restrictions imposed by the *Veterinary and Para-Veterinary Professions Act* are especially onerous in a field situation where researchers cannot predict when they are going to need the services of a veterinarian, and this can lead to significant delays and cost implications, or the work being cancelled altogether.

References

1. South African Department of Science and Technology (DST). White Paper on Science, Technology and Innovation: Science, technology and innovation enabling inclusive and sustainable South African development in a changing world. Pretoria: DST; 2019. Available from: https://www.dst.gov.za/images/2019/WHITE_PAPER_ON_SCIENCE_AND_TECHNOLOGY_web.pdf
2. South African National Biodiversity Institute, Lewis Foundation. A human capital development strategy for the biodiversity sector 2010–2030 [document on the Internet]. c2010 [cited 2021 Apr 21]. Available from: http://www.greenskills.co.za/wp-content/uploads/2015/07/Biodiversity-Human-Capital-Development-Strategy-20101.pdf
3. Retief J. Opinion: Mboweni's plan – 7 considerations that will actually help SMEs. Fin24. 2019 Sep 28. Available from: https://www.news24.com/fin24/opinion/opinion-mbowenis-plan-7-considerations-that-will-actually-help-smes-20190927
4. Hamer M, Behr K, Engelbrecht I, Richards L. Permit requirements, associated challenges and recommendations for biodiversity collections and research in South Africa. S Afr J Sci. 2021;117(9/10), Art. #11765. https://doi.org/10.17159/sajs.2021/11765
5. South African Department of Environmental Affairs (DEA). National Biodiversity Research & Evidence Strategy (2015–2025): Annual research science-policy interface report 2017/18 [document on the Internet]. No date [cited 2021 Apr 21]. Available from: https://chm.cbd.int/api/v2013/documents/6B9F3AE1-94C5-26A2-CAFE-91CEBD16C040/attachments/National_biodiversity_and_evidence_research_strategy_(NBRES).pdf
6. Kerley GIH, Behrens KG, Carruthers J, Diemont M, Du Plessis J, Minnie L, et al. Building assessment practice and lessons from the scientific assessment on livestock predation in South Africa. S Afr J Sci. 2019;115(5/6), Art. #5766. https://doi.org/10.17159/sajs.2019/5766
7. Alves RJV, Weksler M, Oliveira JA, Buckup PA, Pombal JP Jr., Santana HRG, et al. Brazilian legislation on genetic heritage harms Biodiversity Convention goals and threatens basic biology research and education. Anais da Academia Brasileira de Ciências. 2018;90:1279–1284. http://dx.doi.org/10.1590/0001-3765201820180460
8. Prathapan KD, Pethiyagoda R, Bawa KS, Raven PH, Rajan PD. When the cure kills – CBD limits biodiversity research. Science. 2018;360:1405–1406. https://doi.org/10.1126/science.aat9844
9. Crouch NR, Douwes E, Wolfson MM, Smith GF, Edwards TJ. South Africa's bioprospecting, access and benefit-sharing legislation: Current realities, future complications, and a proposed alternative. S Afr J Sci. 2008;104:355–366.
10. Wynberg R, Laird S. Bioprospecting, access and benefit sharing: Revisiting the 'grand bargain'. In: Wynberg R, Schroeder D, Chennells R, editors. Indigenous peoples, consent and benefit sharing. Dordrecht: Springer; 2009. p. 69–86. https://doi.org/10.1007/978-90-481-3123-5_5
11. Laird S, Wynberg R, Rourke M, Humphries F, Muller MR, Lawson C. Rethink the expansion of access and benefit sharing. Science. 2020;367:1200–1202. https://doi.org/10.1126/science.aba9609
12. Scholz AH, Hillebrand U, Freitag J, Cancio I, dos S Ribeiro C, Haringhuizen G, et al. Finding compromise on ABS & DSI in the CBD: Requirements & policy ideas from a scientific perspective [document on the Internet]. c2020 [cited 2021 Apr 21]. Available from: https://www.dsmz.de/fileadmin/user_upload/Collection_allg/Final_WiLDSI_White_Paper_Oct7_2020.pdf

AUTHORS:
Michelle Hamer[1,2]
Karin Behr[3]
Ian Engelbrecht[1]
Leigh Richards[4]

AFFILIATIONS:
[1]Natural Science Collections Facility, South African National Biodiversity Institute, Pretoria, South Africa
[2]School of Life Sciences, University of KwaZulu-Natal, Pietermaritzburg, South Africa
[3]Foundational Biodiversity Science, South African National Biodiversity Institute, Pretoria, South Africa
[4]Durban Natural Science Museum, Durban, South Africa

CORRESPONDENCE TO:
Michelle Hamer

EMAIL:
M.Hamer@sanbi.org.za

HOW TO CITE:
Hamer M, Behr K, Engelbrecht I, Richards L. Permit requirements, associated challenges and recommendations for biodiversity collections and research in South Africa. S Afr J Sci. 2021;117(9/10), Art. #11765. https://doi.org/10.17159/sajs.2021/11765

ARTICLE INCLUDES:
☐ Peer review
☒ Supplementary material

KEYWORDS:
museums, herbaria, legislation, compliance, biodiversity surveys

PUBLISHED:
29 September 2021

Permit requirements, associated challenges and recommendations for biodiversity collections and research in South Africa

South Africa is frequently cited as being a megadiverse country, with high numbers of species, ecosystems and biomes, and high levels of endemism or uniqueness of species and habitats. A number of globally recognised products such as a detailed national vegetation map[1], a comprehensive strategy for protected areas expansion[2], Red List assessments of plant species[3], mammals, birds, reptiles[4] and butterflies[5]. A comprehensive scientific National Biodiversity Assessment is carried out every 6 or 7 years.[6,7] A large number of books on South African plants, animals and fungi have been published and these books present information not only for scientists but also for the public, environmental impact practitioners, conservationists, the tourism industry, bioprospectors, biotraders and decision-makers. All these biodiversity-related products rely on sound scientific knowledge of species, gathered over decades through expeditions and surveys and research in fields that include taxonomy and biogeography. Material collected and researched is housed in museums and herbaria across the country, and the specimens and associated data are used by scientists and postgraduate students globally on an ongoing basis. Despite the vast amount of knowledge and data, many gaps in the knowledge of South Africa's biodiversity still exist. Over 200 new South African species are described each year, and the application of molecular phylogenetic approaches is revealing a large amount of cryptic diversity in taxa that were considered well known.[8-10] The continued expansion of collections and knowledge is critical for the conservation and sustainable use of species and ecosystems, as well as for understanding the impacts of climate change and other forms of global change on biodiversity.[11,12]

Collection of plants and animals by scientists in South Africa has been regulated through a permitting system for decades. In the last 10 to 15 years, however, the permitting requirements have become increasingly complex and, for most biodiversity scientists, determining what permits are required is daunting. This has a significant impact on the research, capacity development and natural science collection community's mandated work.

Two recent projects have been established through the South African Department of Science and Innovation as part of their South African Research Infrastructure Roadmap (SARIR). The Natural Science Collections Facility is a network of museums, science councils and university herbaria that hold preserved biodiversity specimens and materials. The Biodiversity Biobanks South Africa includes institutions that hold mostly cryopreserved materials such as frozen tissues (e.g. reproductive, blood, muscle), DNA extracts and cultures of microbial organisms. In line with global initiatives for large research infrastructure, the purpose of these SARIR projects is to serve researchers both nationally and internationally through acting as a repository and providing access to materials and data for research and development projects, especially where these are of societal and economic benefit. The challenges faced by participating institutions in terms of permits to collect, accept donations, and house and share materials is a major constraint to the achievement of the objectives of the SARIR initiatives.

In this Commentary, we summarise the current legal requirements for the collection and use of biodiversity material for non-commercial research purposes in South Africa, highlight the main challenges from the perspective of researchers and natural science collection curators and managers, and make recommendations for addressing the challenges identified.

The global context for permits relating to biodiversity

South Africa is a signatory to the Convention on Biological Diversity (CBD), the Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilisation, a supplementary agreement to the CBD, and to the Convention on International Trade in Endangered Species (CITES), which is a multinational treaty that aims to ensure that international trade in wild species of plants and animals does not lead to their extinction. The CBD acknowledges that there are significant gaps in knowledge, and that these need to be addressed in order for the Convention's objectives to be met. The Nagoya Protocol states that signatory countries should establish mechanisms to 'create conditions to promote and encourage research contributing to biodiversity conservation and sustainable use'[13], and CITES makes allowances for preserved material, DNA extracts and tissue samples for analysis of CITES-listed species that is exchanged for scientific research purposes. It is clear that the need to enable research on biodiversity was recognised when these three conventions were developed, but in many countries this intention has not been adequately addressed.

National legislation

Biodiversity samples, specimens and collections can be considered as research infrastructures, biological or genetic resources, agents of disease or heritage assets, depending on which legislation is being considered. The national legislation that gives effect to the CBD, CITES and the Nagoya Protocol is the *National Environmental Management: Biodiversity Act* (NEMBA) and its associated Regulations. In terms of biodiversity research and permits, NEMBA only legislates for those species that are listed on CITES, or on the national Threatened or Protected Species (ToPS) list, or where material will be used or potentially used for commercial projects or where it will be exported from the country for any type of research purpose.

The Threatened or Protected Species Regulations of NEMBA were first published in 2007 but have undergone several amendments. The ToPS list that is currently in use is still the original one published in 2007 and covers

30 mammal, 32 bird, 13 reptile, 2 amphibian, 21 fish, 66 plant and 14 invertebrate species and 10 invertebrate genera. The Regulations and list have been under review for more than a decade.

The challenges for scientific institutions and researchers working on ToPS is the number of different processes that must be followed, and confusion around the authority for these. Registration as a scientific institution and a standing permit are required for institutions, and application forms are available on the South African Department of Forestry, Fisheries and Environment (DFFE)'s website, but the responsible authority is unclear. The issuing authority for ToPS permits and for registration as a scientific institution was originally the provincial authority but this has been amended several times. According to the 2011 amendments, the Minister is the issuing authority for permits for all organs of state and for restricted activities relating to marine ToPS, and in the 2012 amendments, restricted activities in a national protected area were added. According to the Constitution of South Africa, organs of state include institutions 'exercising a public power or performing a public duty in terms of any legislation'. This would cover all national, provincial and municipal museums as well as universities. This suggests that the national Department, and not the provincial authorities, should issue ToPS permits to these institutions, but currently some of the provincial authorities are issuing these permits while others are not. Registration of scientific institutions appears to currently be through DFFE, although this is delegated to the provinces in the Regulations.

ToPS permits cannot be issued retrospectively, but listed species may be inadvertently collected, especially in the case of plants or invertebrates which are difficult to identify in the field. Without the ToPS permit, such specimens cannot legally be deposited in a collection institution and the collector is at risk of prosecution. Experts carrying out surveys also often unexpectedly come across a suspected ToPS, but without a permit, they cannot collect it to confirm the record. The time and expense of returning to a site after a permit has been accessed and the challenges of finding plants that emerge or flower for short periods or are short-lived or small invertebrates mean that this is often not feasible, so valuable new records are lost.

In line with the international Convention, Section 15 of the CITES Regulations deals with 'Scientific exchange', which is intended to facilitate research by not requiring the same application process as commercial traders and hunters. A recent Conference of the Parties (CoP18) notification extended the exemption from normal permitting requirements to DNA extracts or frozen or preserved tissues used for forensics or research. However, scientific institutions need to be registered for CITES, and they need to have what is referred to as a 'label', which is an official form with the CITES acronym, issued by the relevant authority, which accompanies materials being sent out of the country or imported, and which declares the nature of the contents of a parcel and that it is for scientific exchange. Most of the South African collection institutions are CITES registered, but whether they meet the criteria specified internationally (see Supplementary table 1) is uncertain. The process for additional institutions to register is unclear and does not currently seem possible. None of the permit issuing authorities are currently providing CITES 'labels', which means that for scientific exchange, the normal CITES permitting application route has to be followed. For import of materials, the Regulations are unclear, stating that for import of materials of species on Appendix II and III 'prior presentation of either an export permit or a re-export certificate' is required. What is meant by 'presentation' and to whom this refers is unclear.

The Bioprospecting, Access and Benefit Sharing (BABS) Regulations were first gazetted in 2008 but these were extensively amended in 2015. A permit from the relevant provincial department where the material was collected is required to export any biological material from South Africa. This includes specimens and tissue samples or DNA samples that are being donated or supplied to international researchers or institutions, or being sent for analysis, even if the intention is not bioprospecting. While export permits are issued by some of the provincial authorities for research purposes, it appears that these are being issued in terms of the provincial ordinances rather than the national Regulations, and it is uncertain whether the provincial export permits comply with the BABS Regulations and with the Nagoya Protocol requirements.

The Nagoya Protocol states that parties (countries) should issue a permit as evidence that access to genetic resources was based on prior informed consent (e.g. collecting permit, written landowner consent, export permit) and mutually agreed terms (e.g. specified benefits such as access to research results, collaboration with local researcher or student, and material transfer agreement). The recipient of the material (importer) will need an import permit, and this may not be issued by authorities, especially those under the European Union, if the provincial export permit is not explicit about it being issued in terms of the national legislation. An export permit that does not provide evidence of complying with the BABS Regulations may result in the import permit for the recipient being denied, or the material being blocked at a port of entry, or a manuscript resulting from research on the material not being accepted for publication in a European journal.[14]

The Nagoya Protocol also requires that signatories must make information on permits available to the Access and Benefit Sharing Clearing-House, in the form of an 'Internationally Recognised Certificate of Compliance' or IRCC. For South Africa there are 33 IRCCs, but only 4 of these are for non-commercial use. It appears that for most export permits there are no IRCCs issued by provinces, and so international authorities cannot check whether the material for which an import permit has been applied meets the Nagoya Protocol requirements.

A further challenge relates to specimens collected under standard collecting permits or without permits. According to the Nagoya Protocol, prior informed consent to access the materials is required. A collecting permit or written landowner consent and an export permit issued under the BABS Regulations from the relevant authority could be considered as prior informed consent. Without this, specimens or materials cannot be exported for research purposes or for sequencing or other analysis. While there are more stringent requirements for materials collected after the Nagoya Protocol came into effect, i.e. October 2014, this is not addressed in the Regulations.

Specimens or samples in collections cannot be used for any type of research that may be linked in any way to bioprospecting or commercialisation, even if this is in the future and/or downstream (e.g. the specimen may in future be used for chemical analysis for any compound that may have a pharmaceutical potential) because this requires an Exploration Permit, and change in use from the original intent of collection is not allowed, even though the Nagoya Protocol promotes accommodating change of intent.[13] There have been examples in which specimens were collected on an ordinary provincial research permit, and these were later used for biocompound extraction in a study investigating potential for pharmaceuticals. This is illegal according to the BABS Regulations.

The *National Forests Act (No. 84 of 1998)* regulates activities involving protected trees and in state forests, including research. Application forms for licences for such activities are available on the DFFE website, but there are no details about submission of the application forms except on the actual form where it is stated that these must be sent or delivered to the nearest Forestry Office of the DFFE. The website does not include a list of the state forests nor any contact information for queries or for submitting the application form. The application forms do state that the licence does not exempt the applicant from complying with other legislation, which suggests that any activities regulated under NEMBA or provincial ordinances will still require those permits.

The *Animal Diseases Act (No 35 of 1984. Section 20)* is critically important for protecting livestock and people from disease. The South African Department of Agriculture, Land Reform and Rural Development (DALRRD) provides guidelines for the application of a permit under Section 20 of the Act[15], but the permitting process is still opaque and complex (Supplementary tables 1 and 2). Because new diseases emerge, a set list of taxa cannot be provided, and so any animal or microbe may require a Section 20 permit. Handling, storage facilities and protocols at the institution, transport and waste management are

considered in the assessment of applications, especially if there is a perceived risk of disease. If any biobanked materials are held and these do not have permits, detailed records or data, they could be incinerated by DALRRD.

Collection institutions often receive dead animals from the public, for example, birds or small mammals that have been killed by domestic cats or dogs, or animals killed on the road. According to the Act, the institution would need to be certified and meet the requirements above, but, even then, such specimens should probably not be accepted because they were collected without a Section 20 permit.

While the Act has been in force for almost 40 years, it appears that only in the last few years has there been an effort to expand its application to research on all inland taxa, many of which are highly unlikely to be flagged by the State Veterinarians as being associated with the spread of diseases of concern. There is limited knowledge and understanding of the Act amongst researchers, and little understanding of the scope and scale of work that is done by biodiversity researchers by the Animal Disease Control unit of DALRRD. If Section 20 permits are applied as required, it is unlikely that the State Veterinarians and the permitting office will be able to service permit requests within the 3-month period they suggest is required for applications to be processed.

The *Agricultural Pests Act* is relevant for controlling the import or export of pests which pose a risk to South Africa or other countries' agriculture. Import and export permits are required for agricultural products (fruit, vegetables, ornamental plants) and associated packaging, but no permits appear to be required for other biological specimens. While this may be a relief to collection institutions, if herbaria or museums do not have phytosanitary certificates for export or import of preserved herbarium or dried animal material, it is possible that donations and loans of specimens may be stopped by customs and biosecurity officials. A recent article[16] gave two examples of consignments of historical herbarium specimens that were incinerated by Australian biosecurity officials. Some institutions in the USA are also concerned about sending out loans because their own biosecurity agencies may prevent their re-entry into the country because of the risk of importing pests.

In addition to CITES, ToPS and BABS export permits, some specimens may also need export permits if they are considered to be significant South African heritage objects. In terms of the *National Heritage Resources Act (No. 25 of 1999)* and a notice published in April 2019, natural history 'type specimens' require an export permit, whether they are being sent permanently or on loan. There are many forms of types (e.g. syntypes, paratypes, topotypes) and not all of these have the same value in terms of heritage. It is also unclear whether specimens collected outside South Africa but housed in collections here should also be considered as national heritage.

Provincial ordinances for nature conservation

Each province has its own legislation controlling the collection, import and export of plants and animals. This legislation is extremely complex to navigate for collections staff and researchers. For some provinces, the ordinances predate the political transition of 1994 democracy and are applicable to the former homelands. For example, the Eastern Cape currently operates under three outdated ordinances (Transkei, Ciskei and Cape Province), and how these are interpreted is unclear. The North West Province has a similar challenge with its ordinances coming from Bophuthatswana, the Transvaal and the Cape Province. Some provinces have drafted new biodiversity legislation in the form of an Act but these are not yet in force (e.g. KwaZulu-Natal and North West Province).

At the time of writing, most of the provinces do not provide guidelines for permits or contact details for the permitting office or access to the application forms online (Supplementary table 3). The provincial legislation and lists of protected species are also not provided. An exception is CapeNature, which has a webpage providing all relevant information.

Each province has different categories, and lists of protected species and activities relating to these require specific permits. The categories and definitions vary between provinces, and the species included in these differ from the national lists and from each other. In many cases, the names and classifications are outdated and are not standardised across provinces, and the criteria and rationale for species included is unclear. In most cases, the sections of the ordinances relating to animals are in the context of hunting, game ranching and poaching or recreational or commercial fishing, and in the case of plants, commercial harvesting, propagation and sales. In general, provincial permitting therefore does not specifically cover the collection of specimens or samples for scientific purposes, and researchers are subject to the same rules and regulations as hunters and commercial ventures.

In general, provinces require registration of a research project and a collecting permit for protected areas, and a permit to collect listed species outside protected areas and to export these out of the province. Collecting along roads is prohibited in provinces except the Northern Cape and Western Cape, where this can be included in permits. Landowner permission is required for collecting outside of the protected areas under the control of the issuing authority, and this may be a private landowner or a state entity such as SANParks, iSimangaliso Wetland Park Authority, the South African National Biodiversity Institute, a municipality or a state forest. In the case of SANParks, the project must be registered with them and a licence to collect must be issued, but the provincial permit is still required. Possession, donation, accepting a donation, transporting, importing into or exporting from the province of any of the listed species also requires permits.

Discussion

Originally, we intended developing guidelines on permitting for collection institutions, but after studying 25 pieces of legislation, including amendments, covering over 1500 pages and 15 different lists of species, and trying to interpret these in the context of collection institutions and research activities, it became clear that this was an impossible task. The complexity of the current requirements (Supplementary tables 1–3), the lack of access to useful information (Supplementary table 3), unclear or irrational legislation, a series of amendments that require checking back to the original document to try follow changes, published but not promulgated legislation, together with delayed or no response from some authorities to permit requests or queries, leaves institutions struggling to carry out their mandated responsibilities while complying with all legislation. The threat of institutions receiving fines of up to ZAR5 million or individuals being imprisoned for up to 5 years could result in a decline in the expansion of collections and in provision of specimens or samples for research and development. The permitting requirements also affect researchers from outside collection institutions and environmental impact assessment practitioners who collect material that they need identified or deposited in a collection institution. Large quantities of important material may be discarded or destroyed because the material cannot be legally deposited in a collection institution. International collaboration – which results in accelerated knowledge generation, capacity development, access to new technologies and increased research investment – is also on the decline due to these impediments. These impacts are not the intended consequence of any of the legislation, the CBD, CITES or the Nagoya Protocol, which all recognise the need for finding ways of streamlining regulation processes for research.

Recommendations

Globally, the challenges with permits, especially related to the Nagoya Protocol, for natural science collections have been raised[17], but it has been recognised that country-based solutions are needed. Amendment of legislation is usually a lengthy and costly process, and so while changes may be needed, this is not a short-term solution for urgently addressing the permitting challenges.

The national legislation relating to permitting serves several important functions, including meeting the requirements of global conventions that allow South Africa to trade internationally, protecting our economy, protecting populations of rare species, and ensuring that special specimens sent out of the country can be legally retrieved if necessary.

The BABS legislation is becoming critical to protect the interests of local researchers, not only in terms of preventing loss of access to specimens, but also the ability to publish in international journals that are increasingly checking Nagoya Protocol compliance.[14,18] This means that it is not reasonable to expect research or collection institution activities to simply be exempt from all permits. However, it is reasonable to expect that the permitting requirements and processes will be transparent, accessible, responsive and rational. Supplementary table 2 illustrates the level of uncertainty and the complexity associated with permits that confront collection institutions and researchers.

Streamlining processes for permits for research and the other activities of collection institutions could have a significant positive impact. At a provincial level, permitting offices deal with thousands of permit applications each year, mostly related to hunting, trade and harvesting. Their expertise and priority may not be academic research on obscure taxa, and so mechanisms are required to reduce the burden, not only on researchers but also on the issuing authorities. Annual collecting permits issued to institutions and integrated permits that cover a range of activities and taxa would reduce the burden. Annual reports and data sets from field trips can still be submitted for the permits to be renewed and the need to notify managers of protected areas of proposed research is also a reasonable expectation.

The research and collection institutions need to ensure that they play their part in being credible, professional, accountable and ethical, which will increase levels of trust by the issuing authorities. There have been cases of international 'biopiracy' by scientists[18,19], and this can undermine efforts to work with the authorities to streamline permitting processes. The loss of potential revenue through biological samples being sent outside the country without a Material Transfer Agreement that restricts its use for commercialisation, or where material is exported and deposited in overseas institutions, inaccessible to South African researchers, are valid concerns, but this type of activity is not effectively regulated by the current complex permit environment, which may actually inadvertently drive non-compliance.

While reduced activity in biodiversity surveys, collections and research may not seem like a high priority outside of those affected, there are impacts for capacity development, knowledge generation, land-use decision-making and the bioeconomy and so initiatives like the Natural Science Collections Facility and the Biodiversity Biobanks South Africa, as well as professional societies and associations should facilitate engagement with the authorities and actively participate in any opportunity that arises for input into reviews of legislation.

Acknowledgements

The Natural Science Collections Facility and the Biodiversity Biobanks South Africa are funded by the Department of Science and Innovation through the South African Research Infrastructure Roadmap programme. The collections community is thanked for contributing to discussions on permitting, and Krystal Tolley (SANBI) is thanked for commenting on a draft of the manuscript.

Competing interests

We have no competing interests to declare.

References

1. Rutherford MC, Mucina L, Powrie LW. The South African National Vegetation Database: History, development, applications, problems and future. S Afr J Sci. 2012;108(1–2), Art. #269. http://dx.doi.org/10.4102/sajs.v108i1/2.629
2. South African Department of Environmental Affairs (DEA). National Protected Areas Expansion Strategy for South Africa. Pretoria: DEA; 2016.
3. Raimondo D, Von Staden L, Foden W, Victor JE, Helme NA, Turner RC, et al., editors. Red List of South African plants. Strelitzia. Volume 25. Pretoria: South African National Biodiversity Institute; 2009.
4. Bates MF, Branch WR, Bauer AM, Burger M, Marais J, Alexander GJ, et al. Atlas and red list of the reptiles of South Africa, Lesotho and Swaziland. Pretoria: South African National Biodiversity Institute; 2014.
5. Mecenero S, Ball JB, Edge DA, Hamer ML, Henning GA, Krüger M, et al., editors. Conservation assessment of butterflies of South Africa, Lesotho and Swaziland: Red list and atlas. Johannesburg / Cape Town: Saftronics / Animal Demography Unit; 2013.
6. Skowno AL, Poole CJ, Raimondo DC, Sink KJ, Van Deventer H, Van Niekerk L, et al. National biodiversity assessment 2018: The status of South Africa's ecosystems and biodiversity. Synthesis report. Pretoria South African National Biodiversity Institute and Department of Environment, Forestry and Fisheries; 2019.
7. Driver A, Sink KJ, Nel JN, Holness S, Van Niekerk L, Daniels F, et al. National biodiversity assessment 2011: An assessment of South Africa's biodiversity and ecosystems. Synthesis report. Pretoria: South African National Biodiversity Institute and Department of Environmental Affairs; 2012.
8. Goodier SA, Cotterill FP, O'Ryan C, Skelton PH, De Wit MJ. Cryptic diversity of African tigerfish (Genus *Hydrocynus*) reveals palaeogeographic signatures of linked Neogene geotectonic events. PLoS ONE. 2011;6(12), e28775. http://dx.doi.org/10.1371/journal.pone.0028775
9. Tolley KA, Conradie W, Harvey J, Measey J, Blackburn DC. Molecular phylogenetics reveals a complex history underlying cryptic diversity in the bush squeaker frog (*Arthroleptis wahlbergii*) in southern Africa. Afr Zool. 2018;53(3):83–97. http://dx.doi.org/10.1080/15627020.2018.1517608
10. Taylor PJ, Kearney TC, Kerbis Peterhans JC, Baxter RM, Willows-Munro S. Cryptic diversity in forest shrews of the genus *Myosorex* from southern Africa, with the description of a new species and comments on *Myosorex tenuis*. Zool J Linn Soc. 2013;169(4):881–902. http://dx.doi.org/10.1007/s13364-016-0291-z
11. Kharouba HM, Lewthwaite JM, Guralnick R, Kerr JT, Vellend M. Using insect natural history collections to study global change impacts: Challenges and opportunities. Phil Trans Roy Soc B. 2019;374(1763), Art. #20170405. https://doi.org/10.1098/rstb.2017.0405
12. Bradley RD, Bradley LC, Garner HJ, Baker RJ. Assessing the value of natural history collections and addressing issues regarding long-term growth and care. BioScience. 2014;64(12):1150–1158. http://dx.doi.org/10.1093/biosci/biu166
13. Nagoya Protocol on access to genetic resources and the fair and equitable sharing of benefits arising from their utilization to the Convention on Biological Diversity: Text and annex. Montreal: Secretariat of the Convention on Biological Diversity; 2011.
14. Marden E, Abbott RJ, Austerlitz F, Ortiz-Barrientos D, Baucom RS, Bongaerts P, et al. Sharing and reporting benefits from biodiversity research. Mol Ecol. 2021;30(5):1103–1107. http://dx.doi.org/10.1111/mec.15702
15. South African Department of Agriculture, Forestry and Fisheries (DAFF). Guidelines for application for a permit under Section 20 of the Animal Diseases Act (Act No. 35 of 86) [document on the Internet]. c2017 [cited 2020 Dec 12]. Available from: https://www.nda.agric.za/vetweb/Epidemiology/Section20/Guidelines%20Section%2020%20Applicants%20V171%2014%20February%202017.docx.pdf
16. Stokstad E. Botanists fear research slowdown after priceless specimens destroyed at Australian border. Science News. 2017 May 11 [cited 2020 Dec 12]. https://doi.org/10.1126/science.aal1175
17. Watanabe ME. The Nagoya Protocol: Big steps, new problems. BioScience. 2017;67(4):400. https://doi.org/10.1093/biosci/bix019
18. Law Y-H. Illicit centipede raises thorny question: Should journals have refused to publish a paper about it? Science Insider. 2021 February 10. Available from: https://www.sciencemag.org/news/2021/02/illicit-centipede-raises-thorny-question-should-journals-have-refused-publish-paper
19. Fukushima C, West R, Pape T, Penev L, Schulman L, Cardoso P. Wildlife collection for scientific purposes. Conserv Biol. 2021; 35:5–11. https://doi.org/10.1111/cobi.13572

AUTHORS:
Walter Cristiano[1,2]
Cristina Giacoma[2]
Mario Carere[1]
Laura Mancini[1]

AFFILIATIONS:
[1]Unit of Ecosystems and Health, Department of Environment and Health, Italian National Institute of Health, Rome, Italy
[2]Department of Life Sciences and Systems Biology, University of Turin, Turin, Italy

CORRESPONDENCE TO:
Walter Cristiano

EMAIL:
walter.cristiano@guest.iss.it
walter.cristiano@unito.it

DATES:
Received: 15 Feb. 2021
Revised: 25 May 2021
Accepted: 08 June 2021
Published: 29 Sep. 2021

HOW TO CITE:
Cristiano W, Giacoma C, Carere M, Mancini L. Chemical pollution as a driver of biodiversity loss and potential deterioration of ecosystem services in Eastern Africa: A critical review. S Afr J Sci. 2021;117(9/10), Art. #9541. https://doi.org/10.17159/sajs.2021/9541

ARTICLE INCLUDES:
☒ Peer review
☐ Supplementary material

DATA AVAILABILITY:
☐ Open data set
☐ All data included
☐ On request from author(s)
☐ Not available
☒ Not applicable

EDITORS:
Jennifer Fitchett
Yali Woyessa

KEYWORDS:
chemical pollution, biodiversity loss, ecosystem services, African wildlife, ecosystem health

FUNDING:
None

Chemical pollution as a driver of biodiversity loss and potential deterioration of ecosystem services in Eastern Africa: A critical review

Chemical pollution, i.e. the release of anthropogenic chemical substances into the environment, is a driver of biodiversity loss. Although this issue has been widely investigated in high-income countries of temperate regions, there is a lack of data for tropical areas of middle- or low-income countries, such as those in Eastern Africa. Some of the world's richest biomes that are affected by multiple pressures, including chemical pollution, are hosted in this macro-region. However, few studies have addressed the impact of the release of anthropogenic chemical pollutants on the biodiversity, and the related potential implications for the deterioration of ecosystem goods and services in this area. A contribution in systemising the scientific literature related to this topic is, therefore, urgently needed. We reviewed studies published from 2001 to 2021, focusing on the chemical pollution impact on Eastern African wildlife. Despite an extensive literature search, we found only 43 papers according to our survey methods. We focused on wildlife inhabiting terrestrial ecosystems and inland waters. According to our search, Kenya and Uganda are the most represented countries accounting for about half of the total number of reviewed articles. Moreover, 67.4% of the studies focus on inland waters. The spread of anthropogenic chemicals into tropical areas, e.g. Eastern Africa, and their effects on living organisms deserve greater attention in research and politics. We report a weak increasing trend in publishing studies addressing this topic that might bode well. The combined effort of science and governments is crucial in improving the management of chemical pollutants in the environment for achieving the goals of biodiversity conservation.

Significance:

- Chemical pollution represents an underestimated risk for the health of tropical ecosystems in middle- and low-income countries, such as those of Eastern Africa.
- There is a lack of data on chemical pollution effects on wildlife of Eastern African biomes.
- The anthropogenic release of chemical substances affects the health of biodiversity and humans, negatively influences ecosystem services in Eastern Africa, and makes conservation and protection measures less effective.
- There is an urgent need for improving research on chemical pollution effects and promoting a sustainable use of natural resources in Eastern Africa along with better management of farming and mining activities.

Introduction

Biodiversity loss is a global environmental issue. The rate of species extinction has been dramatically increasing over the last few decades.[1-3] This is mainly due to climate change[4] and many anthropogenic pressures that ecosystems face[5]. Amongst these threats, the release and the discharge of chemical pollutants into the environment, i.e. chemical pollution, is a major concern for the health and survival of living organisms, such as plants and animals, including humans.[6,7] Furthermore, such pressure directly affects biodiversity and human health through biomagnification in the food chain.[8] Human well-being is likely to be impaired in a roundabout way on account of ecosystem service damage due to biodiversity loss.[9,10] Ecosystem services represent the benefits provided by nature to humanity, including food, pure water, pharmaceuticals, climate regulation, pollination and pest control, defence against natural hazards, and reduction of infectious diseases.[9,10] For instance, wildlife habitat exploitation and hunting may facilitate the spread of zoonosis.[9] At the same time, anthropogenic pressures such as chemical pollution could lead to loss of biodiversity, and there is evidence that such events might have an adverse impact on basic ecosystem services, e.g. primary production and nutrient recycling, and final ecosystem services, e.g. food and water supply.[7] Therefore, it follows that biodiversity loss and ecosystem services are closely related, and have the potential to affect both human health and global economies.[10]

Because the spread of chemical pollutants in the environment represents a serious concern for ecosystem health[11], holistic approaches such as One Health and Eco-Health have emerged in the last few years. Both these approaches assume that humans, animals, and the ecosystems to which they belong interact in a complex way and at different levels. According to these concepts, the environment, humans, and other animals can affect each other, so protecting ecosystems means protecting ourselves.[12] High-income countries located in temperate areas have adopted measures to mitigate and prevent the release of hazardous chemical substances. For instance, according to the Water Framework Directive, European Union member states are obliged to achieve a good chemical quality status for all their water bodies by including a list of priority chemical pollutants that must be monitored in water, sediment or biota.[13] Unlike temperate ecosystems, where the chemical pollution effects on biodiversity have been widely investigated for decades, only a few studies have addressed the effects of such environmental contamination in tropical areas. These territories often lie in low- and middle-income countries (according to

the classification provided by the United Nations[14]), and host global biodiversity hotspots[15,16]. In tropics, large forests are the dwelling of most of the terrestrial species worldwide in terms of diversity.[17] These ecosystems face the challenges of deforestation and overexploitation of natural resources, which are involved in tropical biodiversity loss together with climate change and invasive species introduction.[18] Agriculture and mining complicate this picture by representing the main drivers of chemical pollution of water and soil. Such activities involve the release of ubiquitous environmental chemical pollutants such as organochlorine compounds (OCs) and heavy metals.[11,19] It follows that chemical pollution is a global emergency concerning the health of all the different ecosystem components, i.e. environment, animals, humans. This is particularly evident in middle- and low-income countries, where the effects of the release of chemical pollutants into the environment are particularly acute. For instance, more than 90% of world pollution-related deaths in children occur in such world regions.[20] In these areas, the adverse effects of climate change may also increase the impacts of anthropogenic pressures on the ecosystem, including chemical pollution.[21,22] This is a worrying scenario because the interactive effect of climate change and the release of chemical pollutants into the environment is rather unknown.[23] Furthermore, the already limited resources and infrastructures of tropical countries may collapse under the burden of environmental hazards exacerbated by climate change.[24] Therefore, further research is needed to address chemical pollution effects in tropical areas of middle- and low-income countries.

Our work aimed to help to systematise the scientific literature on chemical pollution threats, i.e. the release and discharge of anthropogenic contaminants to wildlife in tropical areas, and to provide a benchmark for future precautionary environmental actions. We chose the ecosystems of the eastern African macro-region as a case study because they represent some of the most important world biomes and they are located in tropical middle- and low-income countries.[25] Promoting the need for improving research on the effects of chemical pollution on African tropical ecosystems as a driver for biodiversity loss and ecosystem service degradation is the goal of our review in a wider perspective. Particularly, we focus on the wildlife of inland waters and terrestrial ecosystems, and we discuss potential implications for human health.

Survey method

We summarised the current state of affairs on the effects of anthropogenic chemical pollution on the health of wildlife (i.e. wild vertebrate and invertebrate species) of Eastern African ecosystems. We considered studies that analysed the direct or indirect impacts of chemicals on animal biota. Furthermore, only studies published in English were considered. According to the United Nations classification, there are 18 countries within the Eastern Africa macro-region, excluding the French and British overseas territories (Figure 1).[14] All of these countries are classified as middle- or low-income countries.[26] The discussion is divided into subsections on geographical, social, and political sub-region-specific characteristics.

We collated scientific articles published in accredited journals from 2001 to early 2021 (until the time of submission of this paper) indexed in PubMed, Scopus or Web of Science (all databases) that focused on the environmental contamination of inland waters and terrestrial ecosystems. For our search, we used the country name AND "chemical pollution" AND one of the following keywords at a time: "wildlife", "biodiversity", "ecosystem services". We repeated the query including each of the keywords for each country of Eastern Africa. We then used the country name AND "environmental pollution" and each of the aforementioned keywords and countries. The survey was conducted by searching titles, keywords, and abstracts. We manually checked each result obtained by typing the chosen keywords in combination with the country name and "chemical pollution/environmental contamination" terms, and we included in our review any relevant results, i.e. all the contributions that addressed the topic of our interest in a country of Eastern Africa, and excluded those results not relevant to our topic.

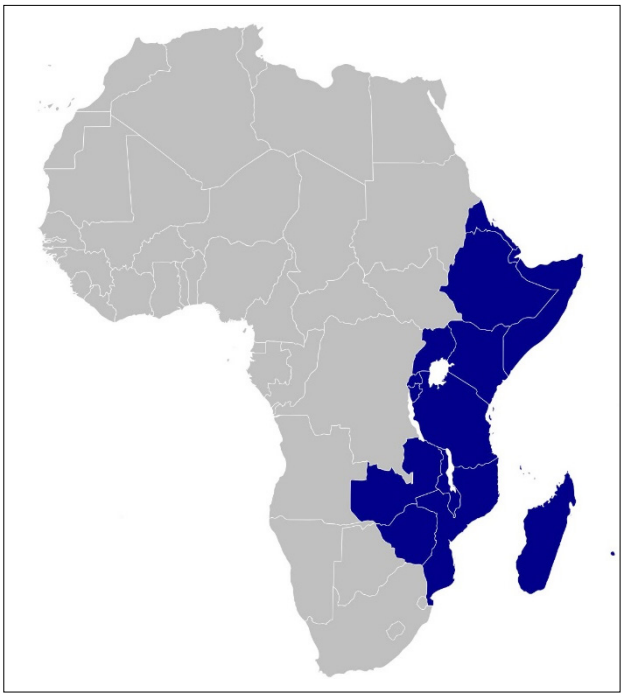


Image: Sannita/Wikimedia Commons/CC-BY-SA-3.0

Figure 1: Map of the African continent. The countries shaded in blue comprise the Eastern Africa macro-region as designated by the United Nations (Statistic Division, Department of Economic and Social Affairs).

We then counted the published articles per year and ran a regression analysis with the aim of identifying potential trends in publishing over time. We did not include the 2021 publications because data for this year are not comparable as a full year was not possible. The regression model was built by setting the year of publication as the predictor variable and the number of articles per year as the dependent variable. The analysis was run in R version 4.0.4. Finally, we calculated the percentages of published articles related to terrestrial and water ecosystems, and the percentage related to each country.

A few references that we found addressed both the direct effects of such pollution sources on biodiversity and the indirect implications for human health and well-being. We discuss and stress such interactions. Due to the limited number of studies, not all the countries within the United Nations classification of Eastern Africa were represented in the search results.

Results and discussions

Eastern Africa is home to some of the most endangered ecosystems on earth, including different biomes[27] such as steppes, savannahs and rainforests. Information on the effects of chemical pollution on sub-Saharan ecosystems is very limited.[15,28] The scarce financial resources due to disadvantaged economies and the simultaneous presence of several emergencies, e.g. political conflicts[29], make it difficult to implement monitoring strategies aimed at investigating the potential risks of environmental pollution on ecosystem health. Furthermore, legislation on chemical pollution is lacking.[30]

Despite an extensive literature search, we found only 43 published articles covering the direct effects of chemical pollution on animals. All the articles retrieved were published between 2001 and January 2021. According to the reviewed studies, the main risks associated with chemical pollution on wildlife are attributed to persistent organic pollutants, especially OCs and heavy metals. OCs are highly persistent in the environment. Exposure to OCs affects living organisms in different ways, including neurotoxic, immunotoxic, cardiotoxic, genotoxic, reproductive, and developmental effects. Furthermore, some OCs cause cancer in different mammals.[31] As for OCs, heavy metals show high

persistence in the environment and negatively affect living organisms. Such chemical pollutants often occur in the environment as a cocktail of substances whose effects on ecosystems are unknown.[32] However, heavy metal exposure leads to oxidation stress, which may induce several adverse effects such as DNA damage and protein modification.[33] There is a close and complex relationship between chemical pollution, biodiversity and human health, and the deterioration of ecosystem services.[7]

We found a slight increasing trend in published studies on chemical pollution and wildlife among countries of Eastern Africa (Figure 2). Although quite weak, the trend is statistically significant. Regression analysis returned a *p*-value equal to 0.03348. Kenya and Uganda were the most represented countries according to our search methods, together accounting for almost half of the retrieved studies, i.e. 20 of 43 articles (Table 1). No data are reported for South Sudan, Eritrea, Djibouti, Malawi, Mozambique, Comoros, Mauritius, and the Seychelles because we did not find a study for these countries through our search. Most of the studies were carried out in inland waters (67.4%). Terrestrial ecosystems accounted for 23.3%, and only 9.3% of the analysed studies involved both aquatic and land ecosystems (Figure 3).

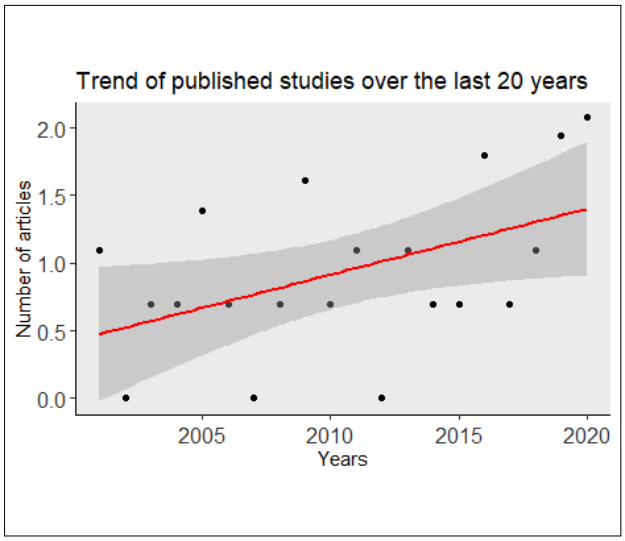


Figure 2: Trend of publications on the chemical pollution effects on the biota of Eastern Africa during the last 20 years.

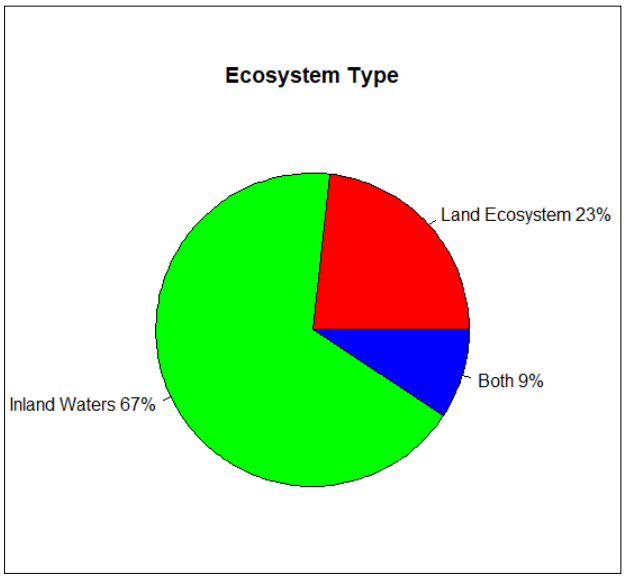


Figure 3: Percentage of articles reviewed according to the target ecosystem investigated.

Table 1: Number of articles retrieved by country in the timespan 2001–2021

Country	Number of published studies
Burundi	1
Ethiopia	4
Kenya	10
Madagascar	3
Rwanda	1
Tanzania	7
Uganda	10
Zambia	8
Zimbabwe	3

Horn of Africa

There are a lack of data for the countries of the Horn of Africa (i.e. Ethiopia, Eritrea, Somalia and Djibouti), although, more recently, a few studies have addressed the effects of effluents on the ecological health of different Ethiopian water basins by assessing macro-invertebrate abundance and diversity. All these investigations found an inverse correlation between water quality and the presence of macro-invertebrates.[34-37]

East African Community

Countries of Eastern Africa that share geographical elements and common interests constitute the East African Community. A batch of studies has addressed the impact of chemical pollution in this area, especially in freshwater ecosystems. In 2019/2020, Musonge et al.[38,39] developed macro-invertebrate-based indexes to investigate the chemical pollution patterns in a biodiversity hotspot in Uganda, the Rwenzori region. Their studies found evidence of pollution according to the different abundances of the macro-invertebrates compared to reference sites. Vertebrate species, such as amphibians, also seem to be sensitive to chemical pollution. In a recent study, in vivo bioassays on a non-target organism, *Xenopus levis* tadpoles, revealed endocrinological, developmental, and behavioural effects due to exposure to water in the Kibale National Park, where 13 pesticides, e.g. carbofuran and 2.4-D amine, were detected. These results indicate a potential health risk for wildlife and humans living in the area.[40]

In the context of freshwater ecosystems, major attention has been placed on Lake Victoria. Bordered by Kenya, Tanzania, and Uganda, this water basin is the largest tropical lake in the world. The environmental chemical contamination of this reservoir is a threat for millions of people living in the surrounding area. The pollution of Lake Victoria also has implications for the economy of the local communities that depend on the exploitation of the water basin for fishing. Therefore, the assessment of environmental pollutants in fish of Lake Victoria is a challenge. At least two studies in Uganda tried to address this issue by studying the impact of pollution on animals inhabiting this ecosystem. In 2019, Badamasi et al.[41] detected biochemical and morphological alterations in three fish species of commercial interest living in this water basin: *Oreochromis niloticus*, *Lates niloticus*, and *Protopterus aethiopicus*. Already in 2006, Focardi et al.[42] reported concentrations of mercury above the limits outlined by the World Health Organization (WHO) in the tissues of common fish species sampled in the Ugandan side of the lake. Evidence of pollution was also found by investigating the impact of agricultural activities on dragonflies. The order Odonata is vulnerable to pesticide use, particularly at the larval stage. In 2009, Martins[43] observed significant differences in the abundance and diversity of dragonflies between pesticide-fished areas and protected areas within the lake. Only a couple of species were found in the impacted areas compared to more than 20 species living in the protected ones.

The pattern of chemical pollution was investigated in Tanzania in four different lakes, including Lake Victoria and Lake Tanganyika. High levels of persistent organic pollutants were assessed in the tissues of fish belonging to the genus *Tilapia*.[44] Such environmental contamination might have direct impacts on human health because these fish represent an important food source for the increasing country population. Lake Tanganyika is also affected by heavy metal contamination. Analyses on fish species, including two important market fish, *Lates microlepis* and *Clarias theodorae*, indicated mercury biomagnification in the tissues exceeding the limit of 0.2 μg Hg/g established by the WHO for vulnerable populations with a high rate of fish consumptions.[45] Mercury biomagnification was also found in Lake Nabugabo, Uganda, by Hanna et al.[46] Different tissues of the Nile perch (*Lates niloticus*) showed mercury contamination, although the reported concentrations were below the value of 0.5 μg/g set by the guidelines of WHO and the Food and Agriculture Organization of the United Nations.[46]

Studies on the water quality of Lake Victoria were also performed in Kenya. Farming activities were suggested as the main drivers for loss of aquatic macro-invertebrates because of pesticide release. Different streams are affected by such disturbance, and this is reflected in the diversity and abundance of many macro-invertebrate species.[47-50] A decreasing gradient of macro-invertebrate richness and diversity was also detected in the Kilombero Valley in Tanzania. This area is a wetland whose streams flowing into the area are commonly used for irrigation. This practice is associated with the degradation of the water quality because of chemical pollutant run-off and leaching.[51] Pesticides may also represent a serious threat to human health, especially via food consumption. Based on the results obtained by investigating the aquatic environment near Morogoro, Groffen et al.[52] have recently pointed out the risk to human health from eating shrimp and fish daily. Indeed, according to the quality guidelines and standard values, exceeding concentrations of heavy metals and persistent organic pollutants – including copper, zinc, and perfluorooctanesulfonic acid – have been found in both invertebrate and vertebrate aquatic species.[52] Mdegela et al.[53] reported evidence of environmental contamination in sewage effluents in Morogoro by using the wild African sharptooth fish (*Clarias gariepinus*) as a model, but only heavy metals were detected. In 2016, Omwenga et al.[54] assessed the impact of pesticides on *Oreochromis niloticus*, a fish commonly used in aquaculture and reared in Kiambu and Machakos Counties, Kenya. They found evidence of different pesticide traces, including dichlorodiphenyltrichloroethane (DDT), i.e. a chemical widely used in sub-Saharan Africa until recent times, on many target fish organs. Nevertheless, the levels of contamination were lower in all the fish sampled than the limits defined by the Food and Agriculture Organization.[54] Exposure to pesticides also has negative implications for birds and mammals. Otieno et al. found evidence of wildlife poisoning due to the overuse of carbofuran[55,56], one of the most toxic carbamate pesticides. By using forensic analyses, they observed that carbofuran and its metabolites were implicated in the poisoning of *Gyps africanus*, an African vulture species.[55,56] Recent findings also showed potential diet-related health risks for a critically endangered primate species, the mountain gorilla (*Gorilla beringei*) living in the Bwindi Impenetrable National Park located in the southwest of Uganda, that feeds on leaves containing concerning DDT levels.[57] Little is known about the exposure to anthropogenic chemical substances of many other primates that face the risk of extinction. In the forests of Uganda, where land expansion critically exposes animals to different pesticides, the impact of cypermethrin, glyphosate, and chlorpyrifos, was assessed in wild chimpanzees (*Pan troglodytes*) and baboons (*Papio anubis*). By acting as endocrine disruptors, such pesticides lead to severe health defects, including facial dysplasia.[58,59] Wang et al.[60] also recently revealed the presence of different pesticides and flame retardants in faecal samples of wild chimpanzees, baboons, red-tailed monkeys (*Cercopithecus ascanius*), and red colobuses (*Piliocolobus tephrosceles*).

Furthermore, a study led by Buhungu et al.[61] in Burundi estimated the ecological status of the Kinyankonge River by applying a Biotic Integrity Index to study the zooplankton communities. The results outlined a degradation of the ecological quality, particularly in the upstream sampling stations.[61] Chemical pollution might also be involved in the micro-evolutionary processes of water organisms. For instance, in 2020, it was suggested by Gomes-Silva et al.[62] that the fitness of invasive wild guppies (*Poecilia reticulata*) living in a river basin of Rwanda may be impaired by local water pollution.

Southeast African countries

Mining activity represents a major threat to the ecosystem health of the Kafue River in Zambia. Syakalima et al.[63,64] reported heavy metal traces, such as copper, zinc, manganese and lead, in samples collected from different fish species and the Kafue lechwe antelope (*Kobus leche kafuensi*) living in the area. Heavy metal concentrations were also found in liver and kidney samples collected from crocodiles (*Crocodylus niloticus*) inhabiting the Kafue River and another Zambian water basin, the Luangwa River.[65] Traces of various pesticides were also detected in the adipose tissue of hippopotami (*Hippopotamus amphibius*) sampled from the Luangwa River.[66] Mining sites were suggested as drivers of bioaccumulation of lead, zinc, copper, cadmium, and cobalt in mammals, as indicated by the analyses carried out in Zambia on the organs of two wild black rat species (*Rattus rattus* and *Rattus tanezumi*). Such exposure to heavy metals involves alterations in gene expression.[67,68] Other studies in the Zambian city of Kabwe showed heavy metal contamination in the muscular system of fish belonging to the *Serranochromis* genus[69], and in liver, lung, stomach, and blood samples obtained from local lizards (*Trachylepis wahlbergii*)[70].

There are very little data accounting for the neighbouring southeast African countries. Impact on macro-invertebrate abundance and diversity was observed in the riverine systems of Bulawayo, Zimbabwe, and these results confirmed the sensitivity of such organisms to urban pollution.[71,72] Similar approaches were applied by Bere et al.[73] in the Manyame River system with the aim of investigating the impact of heavy metals on the ecological health of this site.

African countries in the Indian Ocean

The island of Madagascar is one of the hottest biodiversity hotspots worldwide and hosts a huge number of species, most of them endemic.[74] The International Union for the Conservation of Nature (IUCN) classifies as endangered or critically endangered a large part of the wildlife of Madagascar, including almost all lemur species.[75] The ring-tailed lemur (*Lemur catta*), one of the flagship species of Madagascar, is affected by at least two classes of persistent organic pollutants, i.e. OCs and heavy metals. Different levels of OCs and metals were found in the blood and hair of wild lemurs.[76] The exposure to persistent organic pollutants represents a concern also for another lemur species, the indri (*Indri indri*). Indris living in disturbed forested areas showed higher levels of nickel and cobalt in their serum compared with those living in the pristine forest,[77] although the effects of such metal exposure on the health of this species have still to be fully understood. However, whereas the number of wild indri populations is decreasing, these findings that show a certain level of chemical pressure on the species must be taken into account as they might be important in terms of conservation, especially from a long-term perspective. The hazards of chemical pollution to the wildlife of Madagascar also arise from the use of insecticides, such as Fipronil (class of phenylprazole insecticides), to fight the locust plague to which the country is frequently subjected. Evidence of food chain perturbations due to the control of locusts in termite colonies and their vertebrate predators has been reported.[78]

Final remarks and conclusions

We have presented an overview of published studies on chemical pollution effects on wildlife diversity in Eastern African tropical areas, underlying the indirect effects for human health and ecosystem service impairment. Environmental protection and biodiversity conservation are prerequisites for maintaining good-quality services and benefits provided by ecosystems over time. The evidence that we found in our review, albeit scant, seems to indicate that such equilibrium in many Eastern African countries is potentially affected by chemical pollution. Our outputs may indicate a potential underestimated risk for the wildlife of Eastern Africa, including several critically endangered species. Despite the weak increasing number of studies, research addressing

these issues is still dramatically lacking. Such a knowledge gap may spoil the efforts in biodiversity conservation, and thereby compromise the ecosystem health overall. Indeed, the reduction in biodiversity is likely to impair ecosystem services, which could reduce resources and exacerbate social unrest in middle- and low-income countries.

Understanding and evaluating the effects of chemical pollution on tropical biodiversity and their associated impact on ecosystem services is a challenge.[7] The lack of evidence on this topic in the scientific literature demonstrates the urgent need for an international vision for sub-Saharan environmental studies. Indeed, this scenario strongly affects the chances to address chemical pollution impairment to biodiversity and ecosystem services in these tropical areas, and implement successful environmental protection strategies aimed at preserving the health of African ecosystems as a whole, including humans. A chemical pollution assessment should become an asset in the political programmes of Eastern African states. The establishment of monitoring strategies for priority pollutants, on the basis of the European Union's Water Framework Directive approach[13], could address the application of fit-for-purpose preventive measures in order to protect the ecosystems and to promote human health and safety.

In some countries, such as Uganda and Madagascar, progress is ongoing. The role of the Ugandan Environmental Health Practitioners has been indicated as fundamental in applying the One Health vision to environmental management. Environmental Health Practitioners promote public health by applying the One Health approach to the monitoring of several environmental features, including chemical substances.[79] Similarly, the identification and assessment of chemical pollution has been defined as a priority issue according to the National Health and Pollution Action Plan of Madagascar.[80] These long-term goals represent suitable strategies for developing and implementing solutions to environmental challenges in Eastern Africa. Only a global research effort may meet such goals, and the involvement of different actors, e.g. research institutes, governments and local communities, in achieving these ambitious goals is fundamental. Therefore, high-income countries are being called upon to play their part together with the middle- and low-income countries in overcoming this challenge.

Competing interests

We have no competing interests to declare.

Authors' contributions

W.C.: Conceptualisation, methodology, data collection, data analysis, writing – the initial draft, writing – revisions. C.G.: Validation, student supervision, project leadership. M.C.: Methodology, validation, writing – revisions. L.M.: Conceptualisation, validation, student supervision, project leadership.

References

1. Wake DB, Vredenburg VT. Are we in the midst of the sixth mass extinction? A view from the world of amphibians. Proc Natl Acad Sci USA. 2008;105(1):11466–11473. https://doi.org/10.1073pnas.0801921105
2. Butchart SHM, Walpole M, Collen B, Van Strien A, Scharlemann JPW, Almond REA, et al. Global biodiversity: Indicators of recent declines. Science. 2010;328(5982):1164–1168. https://doi.org/10.1126/science.1187512
3. Ceballos G, Ehrlich PR, Barnosky AD, García A, Pringle RM, Palmer TM. Accelerated modern human-induced species losses: Entering the sixth mass extinction. Sci Adv. 2015;1(5):1–5. https://doi.org/10.1126/sciadv.1400253
4. Soroye P, Newblod T, Kerr J. Climate change contributes to widespread declines among bumble bees across continents. Science. 2020;367(6478):685–688. https://doi.org/10.1126/science.aax8591
5. Vanbergen AJ. Threats to an ecosystem service: Pressures on pollinators. Front Ecol Environ. 2013;11(5):251–259. https://doi.org/10.1890/120126
6. Backhaus T, Snape J, Lazorchak J. The impact of chemical pollution on biodiversity and ecosystem services: The need for an improved understanding. Integr Environ Assess Manag. 2012;8(4):575–576. https://doi.org/10.1002/ie am.1353
7. Noyes PD, Lema SC. Forecasting the impacts of chemical pollution and climate change interactions on the health of wildlife. Curr Zool. 2015;61(4):669–689. https://doi.org/10.1093/czoolo/61.4.669
8. Fu L, Lu X, Niu K, Tan J, Chen J. Bioaccumulation and human health implications of essential and toxic metals in freshwater products of Northeast China. Sci Total Environ. 2019;673:768–776. https://doi.org/10.1016/j.scitotenv.2019.04.099
9. Myers SS, Gaffikin L, Golden CD, Ostfeld RS, Redford KH, Ricketts TH, et al. Human health impacts of ecosystem alteration. Proc Natl Acad Sci USA. 2013;110(47):18753–18760. https://doi.org/10.1073/pnas.1218656110
10. IPBES Global assessment report on biodiversity and ecosystem services. Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services [document on the Internet]. c2019 [cited 2021 Feb 15]. Available from: https://www.ipbes.net/sites/default/files/downloads/spm_unedited_advance_for_posting_htn.pdf
11. Bayen S, Segovia Estrada E, Zhang H, Lee WK, Juhel G, Smedes F, et al. Partitioning and bioaccumulation of legacy and emerging hydrophobic organic chemicals in mangrove ecosystems. Environ Sci Technol. 2019;53:2549–2558. http://dx.doi.org/10.1021/acs.est.8b06122
12. Harrison S, Kivuti-Bitok L, Macmillan A, Priest P. EcoHealth and One Health: A theory-focused review in response to calls for convergence. Environ Int. 2019;132, Art. #105058. https://doi.org/10.1016/j.envint.2019.105058
13. European Commission. Directive 2000/60/EC of the European Parliament and of the Council, 08 12, Official Journal (OJ L 226) [webpage on the Internet]. c2013 [cited 2021 Feb 15]. Available from: https://eur-lex.europa.eu/eli/dir/2013/39/oj
14. United Nations. Country classification [document on the Internet]. c2014 [cited 2021 Feb 15]. Available from: https://www.un.org/en/development/desa/policy/wesp/wesp_current/2014wesp_country_classification.pdf
15. Weiss FT, Leuzinger M, Zurbrügg C, Eggen RIL. Chemical pollution in low- and middle-income countries. Zürich: Swiss Federal Institute of Aquatic Science and Technology; 2016. Available from: https://www.eawag.ch/fileadmin/Domain1/Abteilungen/sandec/publikationen/Chemical_Pollution/Lamics-WEB.pdf
16. Gunnarsson JS, Castillo LE. Ecotoxicology in tropical regions. Environ Sci Poll Res. 2018;25:13203–13206. https://doi.org/10.1007/s11356-018-1887-4
17. Giam X. Global biodiversity loss from tropical deforestation. Proc Natl Acad Sci USA. 2017;114(23):5775–5777. https://doi.org/10.1073/pnas.1706264114
18. Morris RJ. Anthropogenic impacts on tropical forest biodiversity: A network structure and ecosystem functioning perspective. Phil Trans Roy Soc B. 2010;365:3709–3718. https://doi.org/10.1098/rstb.2010.0273
19. Tornero V, Sylvina TJ, Wells RS, Singh J. Eco-toxicants: A growing global threat. In: Yamagiwa J, Karczmarski L, editors. Primates and cetaceans. Tokyo: Springer; 2014. p. 309–332. https://doi.org/10.1007/978-4-431-54523-1_16
20. Landrigan PJ, Fuller R, Fisher S, Suk WA, Sly P, Chiles TC, et al. Pollution and children's health. Sci Total Environ. 2019;650(29):2389–2394. https://doi.org/10.1016/j.scitotenv.2018.09.375
21. Schiedek D, Sundelin B, Readman JW, Macdonald RW. Interactions between climate change and contaminants. Mar Pollut Bull. 2007;54(12):1846–1856. https://doi.org/10.1016/j.marpolbul.2007.09.020
22. Linares C, Martinez GS, Kendrovski V, Diaz J. A new integrative perspective on early warning systems for health in the context of climate change. Environ Res. 2020;187, Art. #109623. https://doi.org/10.1016/j.envres.2020.109623
23. Arenas-Sánchez A, Rico A, Vighi M. Effects of water scarcity and chemical pollution in aquatic ecosystems: State of the art. Sci Total Environ. 2016;572:390–403. https://doi.org/10.1016/j.scitotenv.2016.07.211
24. Fang SC, Rodrigues EG, Christiani DC. Environmental health hazards in the tropics. In: Ryan ET, Hill DR, Solomon T, Aronson NE, Endy TP, editors. Hunter's tropical medicine and emerging infectious diseases. Elsevier; 2020. p. 200–208. https://doi.org/10.1016/B978-0-323-55512-8.00026-0
25. Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J. Biodiversity hotspots for conservation priorities. Nature. 2000;403(6772):853–858. https://doi.org/10.1038/35002501

26. United Nations Department of Economic and Social Affairs, Statistic Division. Methodology [webpage on the Internet]. No date [cited 2021 Feb 15]. Available from: https://unstats.un.org/unsd/methodology/m49/

27. Mittermeier RA, Turner WR, Larsen FW, Brooks TM, Gascon C. Global biodiversity conservation: The critical role of hotspots. In: Zachos FE, Habel JC, editors. Biodiversity hotspots. Berlin: Springer; 2011. p. 3–22. https://doi.org/10.1007/978-3-642-20992-5_1

28. Kandie FJ, Krauss M, Beckers LM, Massei R, Fillinger U, Becker J, et al. Occurrence and risk assessment of organic micropollutants in freshwater systems within the Lake Victoria South Basin, Kenya. Sci Total Environ. 2020;714:1–13. https://doi.org/10.1016/j.scitotenv.2020.136748

29. Aremu JO. Conflicts in Africa: Meaning, causes, impact and solution. African Res Rev. 2010;4(4):549–560. https://doi.org/10.4314/afrrev.v4i4.69251

30. Kumie A, Kloos H. Occupational health and industrial pollution. In: Berhane Y, Haile Mariam D, Kloos H, editors. The epidemiology and ecology of health and disease in Ethiopia. Addis Ababa: Shama Books; 2006. p. 171–195.

31. Jayaraj R, Megha P, Sreedev P. Organochlorine pesticides, their toxic effects on living organisms and their fate in the environment. Interdiscip Toxicol. 2016;9(3–4):90–100. https://doi.org/10.1515/intox-2016-0012

32. Brack W, Aissa SA, Backhaus T, Dulio V, Escher BI, Faust M, et al. Effect-based methods are key. The European Collaborative Project SOLUTIONS recommends integrating effect-based methods for diagnosis and monitoring of water quality. Environ Sci Eur. 2019;31(19):1–6. https://doi.org/10.1186/s12302-019-0192-2

33. Wu X, Cobbina SJ, Mao G, Xu H, Zhang Z, Yang L. A review of toxicity and mechanisms of individual and mixtures of heavy metals in the environment. Environ Sci Pollut Res. 2016;23:8244–8259. https://doi.org/10.1007/s11356-016-6333-x

34. De Troyer N, Mereta ST, Goethals PL, Boets P. Water quality assessment of streams and wetlands in a fast growing East African city. Water. 2016;8(4), Art. #123. https://doi.org/10.3390/w8040123

35. Mengistie E, Ambelu A, Van Gerven T, Smets I. Impact of tannery effluent on the self-purification capacity and biodiversity level of a river. Bull Environ Contam Toxicol. 2016;96(3):369–375. https://doi.org/10.1007/s00128-016-1735-5

36. Desalegne SA. Macroinvertebrate-based bioassessment of rivers in Addis Ababa, Ethiopia. Afr J Ecol. 2018;56(2):262–271. https://doi.org/10.1111/aje.12444

37. Mereta ST, Ambelu A, Ermias A, Abdie Y, Moges M, Haddis A, et al. Effects of untreated industrial effluents on water quality and benthic macroinvertebrate assemblages of Lake Hawassa and its tributaries, Southern Ethiopia. Afr J Aquat Sci. 2020;45(3):285–295. https://doi.org/10.2989/16085914.2019.1671166

38. Musonge PLS, Boets P, Lock K, Ambarita NMD, Forio MAE, Verschuren D, et al. Baseline assessment of benthic macroinvertebrate community structure and ecological water quality in Rwenzori rivers (Albertine rift valley, Uganda) using biotic-index tools. Limnologica. 2019;75:1–10. https://doi.org/10.1016/j.limno.2018.12.001

39. Musonge PLS, Boets P, Lock K, Ambarita MND, Forio MAE, Goethals PLM. Rwenzori score (RS): A benthic macroinvertebrate index for biomonitoring rivers and streams in the Rwenzori Region, Uganda. Sustainability. 2020;12(24):1–18. https://doi.org/10.3390/su122410473

40. Spirhanzlova P, Fini J-B, Demeneix B, Lardy-Fontan S, Vaslin-Reimann S, Lalere B, et al. Composition and endocrine effects of water collected in the Kibale National Park in Uganda. Environ Pollut. 2019;251:460–468. https://doi.org/10.1016/j.envpol.2019.05.006

41. Badamasi I, Odong R, Masembe C. Implications of increasing pollution levels on commercially important fishes in Lake Victoria. J Great Lakes Res. 2019;45(6):1274–1289. https://doi.org/10.1016/j.jglr.2019.09.024

42. Focardi S, Corsi I, Mazzuoli S, Vignoli L, Loiselle SA, Focardi S. Integrating remote sensing approach with pollution monitoring tools for aquatic ecosystem risk assessment and management: A case study of Lake Victoria (Uganda). Environ Monit Assess. 2006;122:275–287. https://doi.org/10.1007/s10661-005-9180-7

43. Martins DJ. Differences in odonata abundance and diversity in pesticide-fished, traditionally-fished and protected areas in Lake Victoria, Eastern Africa (Anisoptera). Odonatologica. 2009;38(3):261–265.

44. Polder A, Müller MB, Lyche JL, Mdegela RH, Nonga HE, Mabiki FP, et al. Levels and patterns of persistent organic pollutants (POPs) in tilapia (*Oreochromis* sp.) from four different lakes in Tanzania: Geographical differences and implications for human health. Sci Total Environ. 2014;488–489:252–260. https://doi.org/10.1016/j.scitotenv.2014.04.085

45. Campbell L, Verburg P, Dixon DG, Hecky RE. Mercury biomagnification in the food web of Lake Tanganyika (Tanzania, East Africa). Sci Total Environ. 2008;402(2–3):184–191. https://doi.org/10.1016/j.scitotenv.2008.04.017

46. Hanna DEL, Buck DG, Chapman LJ. Effects of habitat on mercury concentrations in fish: A case study of Nile perch (*Lates niloticus*) in Lake Nabugabo, Uganda. Ecotoxicology. 2016;25:178–191. https://doi.org/10.1007/s10646-015-1578-6

47. Ndaruga AM, Ndiritu GG, Gichuki NN, Wamicha WN. Impact of water quality on macroinvertebrate assemblages along a tropical stream in Kenya. Afr J Ecol. 2004;42(3):208–216. https://doi.org/10.1111/j.1365-2028.2004.00516.x

48. Masese FO, Muchiri M, Raburu PO. Macroinvertebrate assemblages as biological indicators of water quality in the Moiben River, Kenya. Afr J Aquat Sci. 2009;34(1):15–26. https://doi.org/10.2989/AJAS.2009.34.1.2.727

49. Minaya V, McClain E, Moog O, Moog O, Omengo F, Singer GA. Scale-dependent effects of rural activities on benthic macroinvertebrates and physico-chemical characteristics in headwater streams of the Mara River, Kenya. Ecol Indic. 2013;32:116–122. https://doi.org/10.1016/j.ecolind.2013.03.011

50. Abong'o DA, Wandiga SO, Jumba IO, Van den Brink PJ, Naziriwo BB, Modadi VO, et al. Occurrence, abundance and distribution of benthic macroinvertebrates in the Nyando River catchment, Kenya. Afr J Aquat Sci. 2015;40(4):373–392. https://doi.org/10.2989/16085914.2015.1113397

51. Alavaisha E, Lyon SW, Lindborg R. Assessment of water quality across irrigation schemes: A case study of wetland agriculture impacts in Kilombero Valley, Tanzania. Water. 2019;11(4):1–22. https://doi.org/10.3390/w11040671

52. Groffen T, Rijnders J, Van Doorn L, Jorissen C, De Borger SM, Luttikhuis DO, et al. Preliminary study on the distribution of metals and persistent organic pollutants (POPs), including perfluoroalkylated acids (PFAS), in the aquatic environment near Morogoro, Tanzania, and the potential health risks for humans. Environ Res. 2021;1921–11. https://doi.org/10.1016/j.envres.2020.110299

53. Mdegela RH, Braathen M, Mosha RD, Skaare JU, Sandvik M. Assessment of pollution in sewage ponds using biomarker responses in wild African sharptooth catfish (*Clarias gariepinus*) in Tanzania. Ecotoxicology. 2010;19:722–734. https://doi.org/10.1007/s10646-009-0449-4

54. Omwenga I, Kanja L, Nguta J, Mbaria J, Irungu P. Organochlorine pesticide residues in farmed fish in Machakos and Kiambu counties, Kenya. Cogent Environ Sci. 2016;2(1):1–9. https://doi.org/10.1080/23311843.2016.1153215

55. Otieno PO, Lalah JO, Virani M. Carbofuran and its toxic metabolites provide forensic evidence for furadan exposure in vultures (*Gyps africanus*) in Kenya. Bull Environ Contam Toxicol. 2010;84(5):536–544. https://doi.org/10.1007/s00128-010-9956-5

56. Otieno PO, Lalah JO, Virani M, Jodinko IO, Schramm KW. Carbofuran use and abuse in Kenya: Residues in soils, plants, water courses and the African white-backed vultures (*Gyps africanus*) found dead. Environmentalist. 2011;31(4):382–393. https://doi.org/10.1007/s10669-011-9350-9

57. Amusa C, Rothman J, Odongo S, Matovu H, Ssebugere P, Baranga D, et al. The endangered African great ape: Pesticide residues in soil and plants consumed by mountain gorillas (*Gorilla beringei*) in Bwindi Impenetrable National Park, East Africa. Sci Total Environ. 2021;758(1):1–11. https://doi.org/10.1016/j.scitotenv.2020.143692

58. Krief S, Berny P, Gumisiriza F, Gross R, Demeneix B, Fini JB, et al. Agricultural expansion as risk to endangered wildlife: Pesticide exposure in wild chimpanzees and baboons displaying facial dysplasia. Sci Total Environ. 2017;598:647–656. https://doi.org/10.1016/j.scitotenv.2017.04.113

59. Lacroux C, Guma N, Krief S. Facial dysplasia in wild forest olive baboons (*Papio anubis*) in Sebitoli, Kibale National Park, Uganda: Use of camera traps to detect health defects. J Med Primatol. 2019;48(3):143–153. https://doi.org/10.1111/jmp.12408

60. Wang SR, Steiniche T, Rothman JM, Wrangham RW, Chapman CA, Mutegeki R, et al. Faeces are effective biological samples for measuring pesticides and flame retardants in primates. Environ Sci Technol. 2020;54(19):12013–12023. https://doi.org/10.1021/acs.est.0c02500

61. Buhungu S, Sibomana C, Adjahouinou DC, Ntakimazi G, Bonou CA, Montchowui E. Assessment of the ecological status of the Kinyankonge River (Burundi), using a Biotic Integrity Index of zooplankton (BII-zooplankton). Afr J Aquat Sci. 2020;45(4):442–451. https://doi.org/10.2989/16085914.2020.1753647

62. Gomes-Silva G, Cyubhairo E, Wronski T, Riesch R, Apio A, Plath M. Water pollution affects fish community structure and alters evolutionary trajectories of invasive guppies (*Poecilia reticulata*). Sci Total Environ. 2020;15(730):1–14. https://doi.org/10.1016/j.scitotenv.2020.138912

63. Syakalima M, Choongo K, Nakazato Y, Onuma M, Sugimoto C, Tsubota T, et al. An investigation of heavy metal exposure and risks to wildlife in the Kafue Flats of Zambia. J Vet Med Sci. 2001;63(3):315–318. https://doi.org/10.1292/jvms.63.315

64. Syakalima M, Choongo K, Chilonda P, Ahmadu B, Mwase M, Onuma M, et al. Bioaccumulation of lead in wildlife dependent on the contaminated environment of the Kafue Flats. Bull Environ Contam Toxicol. 2001;67:438–445. https://doi.org/10.1007/s001280143

65. Almli B, Mwase M, Sivertsen T, Musonda MM, Flåøyen A. Hepatic and renal concentrations of 10 trace elements in crocodiles (*Crocodylus niloticus*) in the Kafue and Luangwa rivers in Zambia. Sci Total Environ. 2005;337(1–3):75–82. https://doi.org/10.1016/j.scitotenv.2004.06.019

66. Flåøyen A, Polder A, Mwase M, Almli B, Musonda MM. Pesticide residues in adipose tissue from hippopotami (*Hippopotamus amphibius* L) living in and adjacent to the Luangwa River in Zambia. Onderstepoort J Vet Res. 2005;72(2):181–183. https://doi.org/10.4102/ojvr.v72i2.216

67. Nakayama SMM, Ikenaka Y, Hamada K, Muzandu K, Choongo K, Yabe J, et al. Accumulation and biological effects of metals in wild rats in mining areas of Zambia. Environ Monit Assess. 2013;185:4907–4918. https://doi.org/10.1007/s10661-012-2912-6

68. Nakayama SMM, Nakata H, Ikenaka Y, Yabe J, Oroszlany B, Yohannes YB, et al. One year exposure to Cd- and Pb-contaminated soil causes metal accumulation and alteration of global DNA methylation in rats. Environ Pollut. 2019;252:1267–1276. https://doi.org/10.1016/j.envpol.2019.05.038

69. Choongo KC, Syakalima MS, Mwase M. Coefficient of condition in relation to copper levels in muscle of serranochromis fish and sediment from the Kafue River, Zambia. Bull Environ Contam Toxicol. 2005;75:645–651. https://doi.org/10.1007/s00128-005-0801-1

70. Doya R, Nakayama SMM, Nakata H, Toyomaki H, Yabe J, Muzandu K, et al. Land use in habitats affects metal concentrations in wild lizards around a former lead mining site. Environ Sci Technol. 2020;54(22):14474–14481. https://doi.org/10.1021/acs.est.0c00150

71. Siziba N, Matshisela A, Mwedzi T, Bere T. Macroinvertebrate communities in riverine systems of buffer areas of protected wildland, rangeland and city areas: Implications for conservation of riverine systems on urbanising watersheds. Environ Sci Pollut Res. 2018;25(1):758–770. https://doi.org/10.1007/s11356-017-0487-z

72. Mwedzi T, Siziba N, Odume ON, Nyamazana E, Mabika I. Responses of macroinvertebrate community metrics to urban pollution in semi-arid catchments around the city of Bulawayo, Zimbabwe. 2020;46(4):583–592. https://doi.org/10.17159/wsa/2020.v46.i4.9071

73. Bere T, Dalu T, Mwedzi T. Detecting the impact of heavy metal contaminated sediment on benthic macroinvertebrate communities in tropical streams. Sci Total Environ. 2016;1(572):147–156. http://dx.doi.org/10.1016/j.scitotenv.2016.07.204

74. Ganzhorn JU, Lowry II PP, Schatz GE, Sommer S. The biodiversity of Madagascar: One of the world's hottest hotspots on its way out. Oryx. 2001;35(4):346–348. https://doi.org/10.1046/j.1365-3008.2001.00201.x

75. International Union for Conservation of Nature (IUCN). The IUCN red list of threatened species – Madagascar [webpage on the Internet]. No date [cited 2021 Feb 15]. Available from: https://www.iucnredlist.org/search?landRegions=MG&searchType=species

76. Rainwater TR, Sauther ML, Rainwater KAE, Mills RE, Cuozzo FP, Zhang B, et al. Assessment of organochlorine pesticides and metals in ring-tailed lemurs (*Lemur catta*) at Beza Mahafaly Special Reserve, Madagascar. Am J Primatol. 2009;71:998–1010. https://doi.org/10.1002/ajp.20742

77. Junge RE, Barrett MA, Yoder AD. Effects of anthropogenic disturbance on indri (*Indri indri*) health in Madagascar. Am J Primatol. 2011;73:632–642. https://doi.org/10.1002/ajp.20938

78. Tingle CCD, Rother JA, Dewhurst CF, Lauer S, King WJ. Fipronil: Environmental fate, ecotoxicology, and human health concerns. Rev Environ Contam Toxicol. 2003;176:1–66. https://doi.org/10.1007/978-1-4899-7283-5_1

79. Musoke D, Ndejjo R, Atusingwize E, Halage AA. The role of environmental health in One Health: A Uganda perspective. One Health. 2016;2:157–160. https://doi-org/10.1016/j.onehlt.2016.10.003

80. Madagascar Health and Pollution Action Plan – Accelerating the implementation of actions to reduce pollution-related illness. Global Alliance on Health and Pollution; 2018. Available from: https://gahp.net/hpap-madagascar/

AUTHORS:
Malukhanye S. Mbopha[1,2,3]
Christo Marais[4]
Theo E. Kleynhans[5]
Karen J. Esler[1,3]

AFFILIATIONS:
[1]Conservation Ecology and Entomology, Stellenbosch University, Stellenbosch, South Africa
[2]Ecological Infrastructure Directorate, South African National Biodiversity Institute, Cape Town, South Africa
[3]DSI-NRF Centre for Invasion Biology, Stellenbosch University, Stellenbosch, South Africa
[4]Natural Resource Management Programmes, Department of Environment, Forestry and Fisheries, Cape Town, South Africa
[5]Department of Agricultural Economics, Stellenbosch University, Stellenbosch, South Africa

CORRESPONDENCE TO:
Malukhanye Mbopha

EMAIL:
msmbopha.mbopha@gmail.com

DATES:
Received: 19 July 2020
Revised: 20 Nov. 2020
Accepted: 13 May 2021
Published: 29 Sep. 2021

HOW TO CITE:
Mbopha MS, Marais C, Kleynhans TE, Esler KJ. Unlocking and securing ecological infrastructure investments: The needs and willingness to invest and institutional support mechanisms used. S Afr J Sci. 2021;117(9/10), Art. #8666. https://doi.org/10.17159/sajs.2021/8666

ARTICLE INCLUDES:
☒ Peer review
☒ Supplementary material

DATA AVAILABILITY:
☐ Open data set
☐ All data included
☒ On request from author(s)
☐ Not available
☐ Not applicable

EDITORS:
Bettine van Vuuren
Sydney Moyo

KEYWORDS:
ecological infrastructure, degradation, investments, mechanisms

FUNDING:
South African Department of Environmental Affairs, South African National Biodiversity Institute

Unlocking and securing ecological infrastructure investments: The needs and willingness to invest and institutional support mechanisms used

Ecological infrastructure (EI) is a natural and near-natural functioning ecosystem that delivers a range of essential services to humankind. Examples include mountain catchments, wetlands, coastal dunes, and riparian corridors. In a world where EI is underinvested, rapid degradation and threats such as unsustainable veld-fire regimes, droughts, climate change, and invasive alien plants persist in dominating the ecological landscape. In South Africa, there are government programmes that encourage the restoration, rehabilitation and protection of EI. However, inadequate funding allocations constrain scaling-up and thus necessitate the unlocking of public and private sector investments to augment resources for ecosystem-based management interventions. A systematic literature review was conducted at a global scale to (1) understand the drivers behind EI investments, (2) understand the willingness and desire of private landowners and land users to participate and contribute to EI investments and (3) identify institutional support mechanisms used to encourage investments. Results suggest that the need to invest is driven by growing degradation of EI and the urgency to meet environmental sustainability goals. The willingness to invest is stimulated by the use of economic-based policies and compensatory mechanisms. Public–private partnerships, public policy, and market-based conservation instruments are institutional arrangements executed to protect EI. These include processes and systems used by the institutions to legislate and manage interventions towards fulfilling the conservation objective. Our review contributes to the EI investment research agenda by recommending coordinated efforts to encourage EI investment from both public and private partners. These measures will help to secure financial resources and mobilise investments beyond monetary terms by coordinating planning and developing capacity and reform policies.

Significance:

- Reviewing international experiences on ecological infrastructure investments will help to inform the Natural Resources Management programmes' efforts to upscale the investments essential to conserve natural ecosystems. The lessons from the systematic review will further reveal other related natural ecosystem investment processes from which to learn. Therefore, gaining a global understanding of these lessons provides evidence-based advice for policy development and decision-making processes which seek to protect natural ecosystems for present and future generations.

Introduction

South Africa is biologically diverse[1,2] with a wide range of important ecosystems that deliver essential services to support humankind. Concerns about the rapid rate of environmental degradation and potential effects on society have triggered a need for substantial investment efforts to counteract such impacts whilst advancing sustainability and the National Development Plan agenda for South Africa.[3] The South African government instituted Natural Resources Management (NRM) programmes to maintain and repair ecological infrastructure while alleviating poverty.

The term 'ecological infrastructure' (EI) refers to a natural or near-natural functioning ecosystem that delivers essential services to humankind. Examples include mountain catchments, wetlands, coastal dunes and riparian corridors.[3,4] The ecosystem services (ES) supplied by the EI are equivalent to socio-economic services (e.g. electricity supply) which are derived from built infrastructure. The EI concept carries an economics and development argument essential to attract public attention for ES support and public and conservation policy recognition.[5,6] The NRM programmes invest in EI rehabilitation and maintenance projects to enhance ES delivery and protect biodiversity whilst empowering underprivileged locals.[7,8] In 2012, the demand for financing nationwide ecosystem-based management interventions was estimated to be six times more than available resources[9] and constant and long-term support for interventions is a costly exercise.

The South African conservation legislative framework requires private landowners to be primarily responsible for the protection and conservation of EI on their private land.[10] Despite this legal requirement, the contribution of private landowners towards public good investment is still insignificant.[11] Acknowledging this gap, a well-subsidised and theoretically reputable investment model known as agri-environment schemes[12,13] was chosen as a comparable model to extract lessons essential for NRM guidance on policy advice and operational matters. Agri-environment schemes (AES) are major sources of nature conservation funding to respond to agricultural impacts threatening species, ecosystems and ES delivery in Europe, Australia and the USA.[14-16] This study was motivated by the realisation that NRM's insufficient funding is linked to the reliance on one funding source and absence of investors in the EI discourse.[11,17,18] Increasing concerns posed by unsustainable veld-fire regimes, droughts,

and floods, exacerbated by invasive alien plants, densification of woody pioneer species (bush encroachment), the loss of vegetation cover in some areas and climate change, have brought to the fore the importance of intact EI and the delivery of ES. Degradation of EI and its negative economic impacts on ES in South Africa have further stimulated interest to understand the possibilities of supplementary funding streams. This study reviews (1) the developmental needs or drivers necessitating EI investments, (2) the willingness of private landowners to participate and/or contribute towards EI maintenance and restoration measures and (3) the role of government support, policy regime and institutional arrangements to stimulate cooperation and shared responsibility for EI protection. The purpose of the study was to provide evidence-based conservation policy advice, lessons and insights to inform decision-makers, scientists, policymakers and NRM practitioners.

Methods

We chose a systematic literature review research methodology to gather scientific and non-scientific information[19] to address the objectives of this study. We developed a systematic review protocol to define the review objectives, questions, criteria for source inclusion and exclusion, and keywords (see supplementary material). We created search terms to extract sources from peer-reviewed (Web of Science, Scopus, Science Direct) and grey literature databases (Google and filter bubble) published from 1970 to 2019 (Table 1). The search strategy was then broadened to identify, evaluate and summarise all eligible sources. The search terms were entered in combination with "Agri-Environment Schemes" to retrieve AES studies relevant to the NRM investment model. Truncated words, synonyms, alternative spellings, Boolean logic, and wildcards were used in the search strings.

A pilot search was conducted prior to the definite search to improve the search strategy. To minimise non-target articles, journals of less relevance to the searched topic were excluded after scanning through their titles and abstracts. Search strings were customised based on different database specifications. The search was supplemented by a snowballing approach based on article references. Inductive (concepts emerging from the review process) and deductive (preconceived review concepts) coding methods were employed to extract thematic information from the literature. A PRISMA (preferred reporting items for systematic reviews and meta-analyses) workflow was adopted to show the selection and assessment of collected sources (Figure 1). This phase focussed on reading the article title, keywords and abstract. Subsequently, duplicates were excluded and the remaining sources were read diagonally (introduction, tables, figures, and conclusion) or in entirety depending on relevance. These sources were then imported into ATLAS.*ti* 8.4 (2018) for coding and qualitative analysis. The frequency of mentions from papers were scored and papers were counted as units.

Table 1: Combination of search strings used to compile literature based on three theme areas

Needs and drivers for ecological infrastructure investment	Developmental willingness and desire of private landowners to invest in the maintenance and restoration of ecological infrastructure	Assessment of institutional support and policy mechanisms used to encourage the restoration and maintenance of ecological infrastructure
"ecolog* infrastructur*" OR	"ecolog* infrastructur*" OR	"ecolog* infrastructur*" OR
"ecosystem* infrastructur*" OR	"ecosystem* infrastructur*" OR	"ecosystem* infrastructur*" OR
"environment* infrastructur*" OR	"environment* infrastructur*" OR	"environment* infrastructur*" OR
"soft infrastructur*" OR	"soft infrastructur*" OR	"soft infrastructur*" OR
"natur* infrastructur*" OR	"natur* infrastructur*" OR	"natur* infrastructur*" OR
"green infrastructur*" OR	"green infrastructur*" OR	"green infrastructur*" OR
"natur* capital"	"natur* capital"	"natur* capital"
NOT	NOT	NOT
"biolog* infrastructur*"	"biolog* infrastructur*"	"biolog* infrastructur*"
AND	AND	AND
"invest*" OR	"means*" OR	"polic*" OR
"financ*" OR	"capacit*" OR	"polic* tool*" OR
"fund*" OR	"willing*" OR	"polic* instrument*" OR
"sponsor*" OR	"contribut*" OR	"polic* framework*" OR
"develop*" NOT	"resourc*" OR	"legal framework*" OR
"investigat*"	"support*" OR	"polic* mechanism*"
AND	"develop*"	AND
"need*" OR	AND	"partnership*" OR
"driv*" OR	"partner*" OR	"collaborat*" OR
"caus*" OR	"privat* landowner*" OR	"co?perat*" OR
"motiv*"	"landown*" OR	"cooperat*" OR
AND	"landhold*" OR	"institution* support*"
"Agri-Environment Schemes"	"farmland*" OR	AND
	"farmer*" OR	"Agri-Environment Schemes"
	"stakeholder*"	
	AND	
	"Agri-Environment Schemes"	

*Note: * shows the use of wildcards or truncated words to retrieve alternative word endings.*

The limitation of our methodology is that we cannot entirely guarantee the inclusion of every relevant study due to different languages and unavailability of full-text resources of some sources.

In total, 751 sources were retrieved from databases and Google. Many of the sources that were used were obtained from relevant conservation journals. In the end, 152 sources were distilled.

Results

Needs and drivers for investing in EI

The results show that various drivers necessitate governments and private sector (private commercial companies and non-profit organisations and occasionally civil society members) to take initiative to invest in EI. The motives for investments varied (Table 2, listed in ranks) depending on investor type and tenacious natural or anthropogenic pressures, for example, natural disasters induced by global climate change and agricultural intensification (land-use and land-cover change). The need to protect and conserve biodiversity and to mitigate the effects of agricultural intensification are amongst the biggest drivers of investments in EI.[20] Investments are mainly driven by government for sustainability and by the private sector for social responsibility. Livelihood enhancement through ES (provisioning, regulatory cultural and support services) delivery while sustaining the EI is also an instrumental driver of investment. Other drivers included adapting to and mitigating the effects of climate change and natural disasters and to decrease their severity, and water resource protection.[21] Government remains the leading investor to address all the EI investment drivers/needs categories.

Willingness to invest in EI

The importance of understanding the desire and willingness of private landowners to adopt conservation practices, and participate in and contribute to EI investments is well recognised.[22] This subsection reviews the willingness of private landowners in conjunction with public institutions to accept investment responsibility. Factors associated with willingness determinants and how they relate to each other were also assessed.

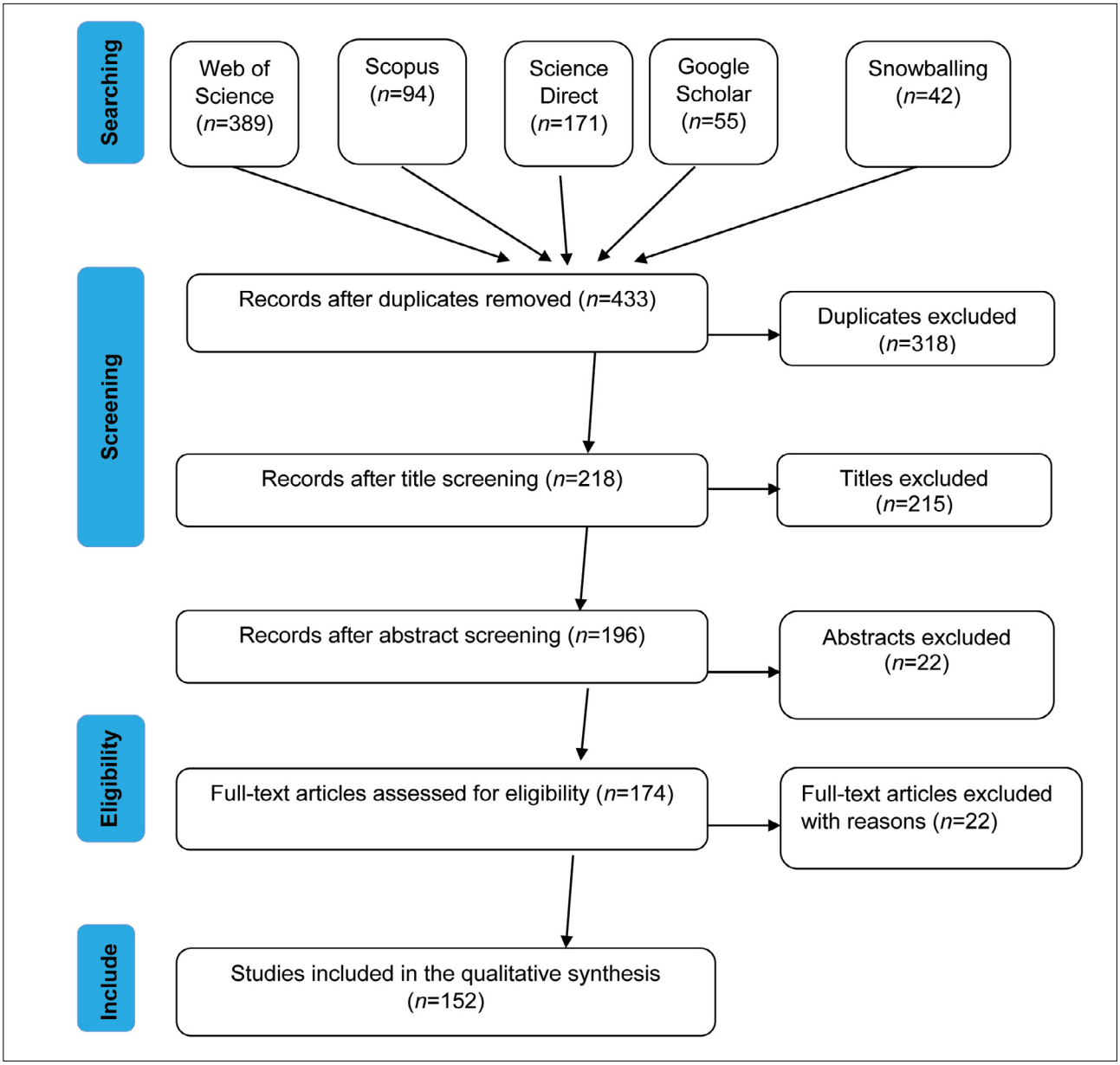


Figure 1: PRISMA diagram sketching the results of articles at searching and screening phases from different databases and sources.

Table 2: The needs or drivers that necessitate investments in ecological infrastructure by different investors. Categories of needs and drivers were derived from ATLAS.*ti* coding of $n=152$ sources reviewed.

Needs or drivers category	Ecological infrastructure investment needs/drivers	Frequency of mention (%)	Ecological infrastructure types	Top-ranking investors per resource and support	Nature of pressure triggering the needs
Biodiversity protection	Enhancement and conservation of biodiversity Eradication of invasive species Protection of biodiversity on private land Habitat protection and maintenance Restoration of endangered ecological communities	27%	Wetlands Forests Endangered species Rivers Ecological communities	Government Private Sector	Natural and anthropogenic pressure
Agricultural or agronomics	Mitigation of detrimental impacts on terrestrial and freshwater ecosystems due to agricultural activities Addressing the impact of land-use and cover change Mitigation of soil erosion	25%	Terrestrial ecosystems Freshwater ecosystems Ecological landscapes	Government Private Sector	Anthropogenic pressure
Ecosystem services and human well-being	Protection of ecosystem services delivery (provisioning, regulatory cultural and support)	20%	Terrestrial ecosystems Wetlands Biodiversity corridors	Government Private Sector	Natural and anthropogenic pressure
Climate change and natural disasters	Lessen drought caused by climate change scenarios Reduction of greenhouse gases Minimise flood risk Carbon sequestration Wildfire risks and extreme disruptive events	14%	Forests Wetlands Landscape/terrestrial ecosystems	Government Private Sector	Natural and anthropogenic pressure
Water management	Address storm-water challenges Improvement of water quality and quantity Protection of water catchments for biodiversity and human well-being	14%	Rivers Catchments Wetlands	Government Private Sector	Natural and anthropogenic pressure

The emerging review themes and frequency of mentions indicate that private landowners' enthusiasm is influenced by five main determinants and the power of the relationship amongst these determinants.

Funding provided and compensation measures encourage private landowners to participate and implement EI rehabilitation and maintenance measures on their land.[23-25] Compensation measures are delivered through the application of market-based conservation instruments, such as economic-based policies that incentivise participating private landowners to deliver ES and comply with the legislation (Figure 2). Socio-economic characteristics and advantages including land ownership, larger farm size, access to information, public awareness and advocacy were stimulating willingness to cooperate towards EI investment, particularly in developed countries. Private landowners who are environmentally conscious, older, experienced in conservation and have a tertiary level education are more likely to participate and steward conservation (Figure 2: Conservation ethics, values and philosophy). Sensitive natural environments and distinctive features trigger conservation interest; Figure 2 shows that where there were unique environmental features such as endangered ecological communities, wetlands and marine assets, many private landowners were keen to counteract degradation through investments and protecting a natural asset they value or rely on. Conservation activism demonstrated by civil society members and non-governmental organisations also placed pressure on authorities and implicated landholders to make contributions towards investments.

Institutional support mechanisms to encourage EI investments

(a) Public policy

The review showed that public institutions and private landowners invest in EI through policy implementation, management, political support, and self-directed environmental awareness and support to advocacy by NGOs.[25,26] Generally, public policies emphasise regulatory enforcement and compliance which compel landowners to implement ecosystem-based management interventions (such as rehabilitation and maintenance) to address EI degradation and loss.[27,28] However, in Latin America and countries such as Finland, policymakers are designing conservation strategies which incentivise landowners to execute sustainable land-use practices that deliver ES, minimise environmental risks and maximise socio-economic development benefits.[29-31]

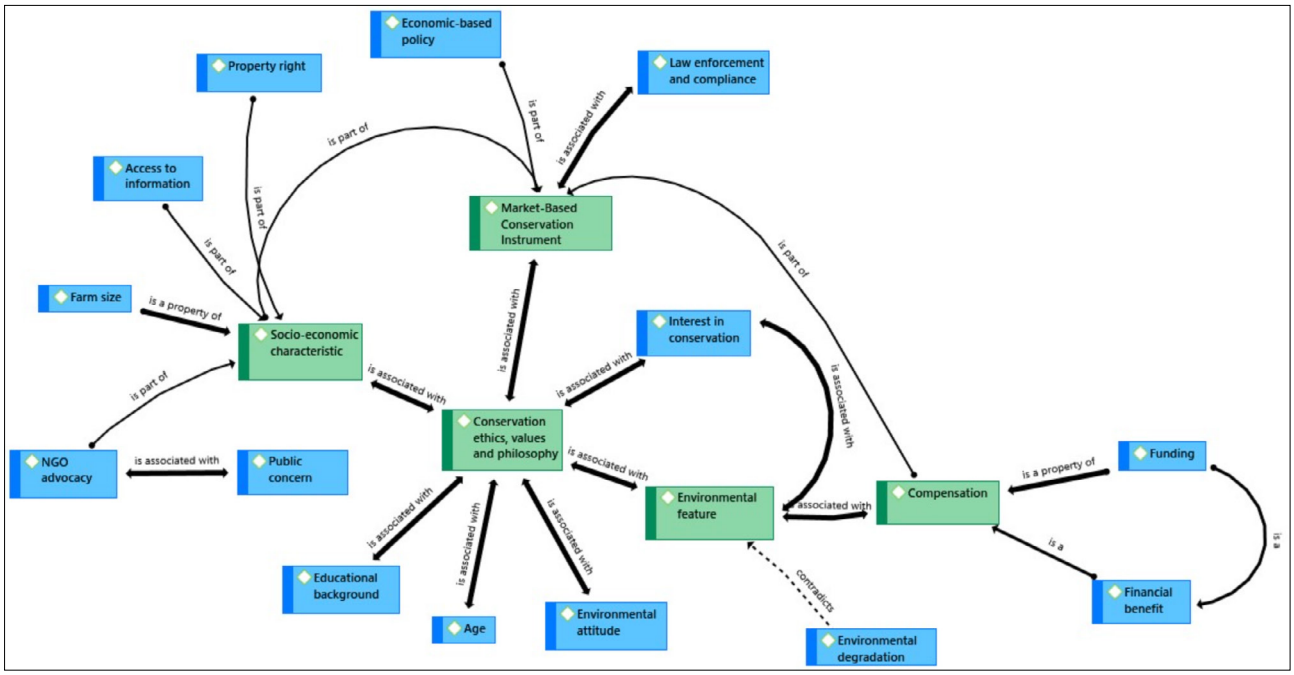


Figure 2: A cognitive map from the review of determinants of willingness to invest in ecological infrastructure by private landowners and other institutions.

(b) Lessons learnt from AES

European Union state members formed incentive programmes to stimulate EI conservation and enhancement on farmlands.[32] These programmes subsidise private landowners who voluntarily adopt and implement agri-environment measures.[16,33] Although AES are widely recognised as major instruments to curb degradation[34], some scholars criticise their efficiency to fully conserve the rural landscape[35,36]. This criticism stems from narrow conservation approaches which focus on individuals or farm-level contracts that are insufficient to achieve wider landscape protection.[37] AES have primarily focussed on incentivising private landowners to comply with schemes' norms and standards (action-based measures). The emerging literature[34,38] suggests that payments should be driven by the desired results (results-based measures)[23].

(c) Public–private partnerships

Public–private partnerships (PPP) are used as cooperative mechanisms to mobilise funding for public goods and services.[39-41] They require effective collaboration between local, provincial and national authorities and government agencies; communities/groups; private landowners; private investors; business sectors; non-governmental organisations; and individual civil society members.[42,43] PPPs have grown rapidly, globally, and have become essential vehicles to encourage private sector investments in conservation.[43,44] However, the current role and commitment of the private sector in long-term arrangements is still insignificant, particularly in developing nations. As a result, PPPs are not mobilising sufficient funding to alleviate EI degradation.[45]

(d) Market-based conservation instruments

Market tools and economic-based mechanisms have been formed to enable ecosystem goods and services delivery, economic growth, and development while protecting EI.[46] These funding vehicles are known as Biodiversity or Conservation Banks.[47,48] They are guided by different policy tools, systems, processes and procedures in different nations. Examples include mitigation biodiversity offsets, mitigation banking, habitat banking, species banking, wetland mitigation, etc.[49,50] They provide compensatory mechanisms through investments to counterbalance degradation and damage resulting from economic practices and developments, and to support interventions that deliver conservation outcomes.[47,51,52] The investments mobilised are used for restoration, maintenance, conservation, and protection of EI and conservation of endangered species.

Discussion

We discuss the implications of the findings on the drivers of investments in EI, the willingness of private landowners to invest in EI and the support mechanisms in place to encourage investments. We present lessons learnt and recommendations to inform and advise natural resources managers and policymakers based on evidence.

Needs and drivers of EI investments

When reviewing the most fundamental drivers or key needs for investments in EI, it was ascertained that conservation, natural disasters, and socio-economic needs put pressure on various institutions and role players to formulate sound measures to restore and rehabilitate (focus on biodiversity and function) and sustain EI. The prevalence of natural disasters and environmental risks stimulate environmental interest because vulnerable landscapes threaten different assets and economic practices. For example, in South Africa, the frequent occurrence of high-intensity and extensive fires, alien plant invasion, and droughts have detrimentally affected EI and economic activities, particularly in the agricultural and forest sectors.[53] The review corroborated the recognition of maintaining and restoring EI as a sustainable technique to abate socio-economic and ecological vulnerabilities and risks.[54] Safeguarding a healthy EI to tactically address climate change related disasters and ecosystem-based adaptation measures have become popular approaches in both ecosystem and livelihood spheres in both developing and developed countries.[55] In its healthy condition, EI contributes to human livelihood improvement and enhances socio-economic development through ES delivery.[56] Drivers of investments in EI can be categorised as 'natural' and 'anthropogenic'; they require both public and private investment attention (Table 2). Therefore, EI maintenance and restoration complement land productivity and economic growth, and safeguard food security. Both private and public institutions are anticipated to devise and implement effective ecosystem-based management interventions, ranging from policies to programmes that mitigate EI degradation.

Willingness and desire to invest in EI

Although the investment contributions could not be quantified, the results revealed that, globally, private landowners are willing to participate and invest in EI when there are compensation measures in place.

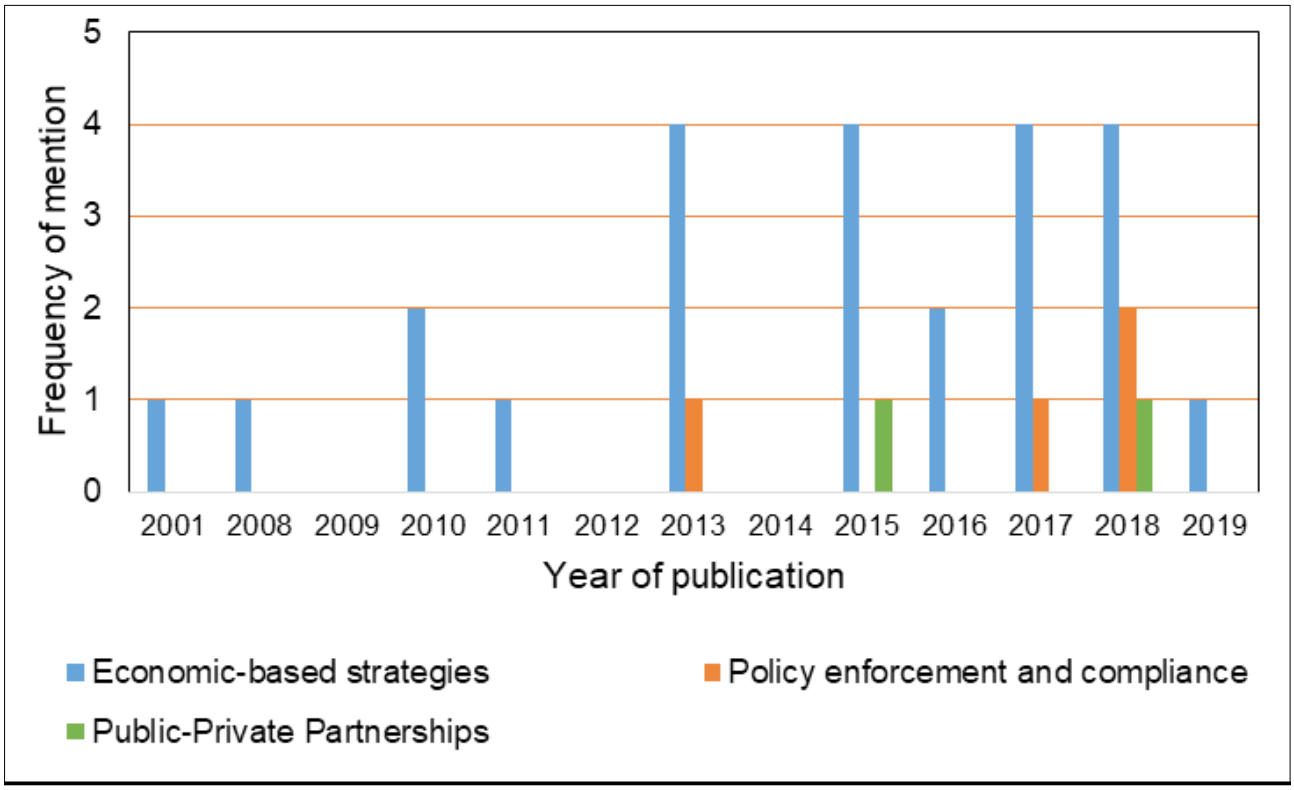


Figure 3: Frequency of mention of economic-based strategies, policy enforcement and compliance, and public–private partnerships in the selected literature over the past two decades.

Financial incentives and funding are generally provided by the government to private landowners who deliver ES through EI management. This analysis suggests that incentive-based policies are pivotal in encouraging the willingness to participate in EI programmes.[57] Beyond awarding economic incentives to attract public and stakeholder participation, there are other critical determinants that define possibilities of stewarding a landscape. Conservation ethics, values and philosophy either encourage or discourage participation and acceptance of the responsibility to restore and maintain EI. Where socio-economic conditions are favourable, private landowners are more likely to conserve EI. These conditions include land rights, land ownership and tenure security; access to information to awaken conservation interest; bigger farm sizes; higher levels of education and communication and advocacy support from both conservation organisations and government; active neighbourhood networks; private landowner conservation consciousness and interest as well as unique and prominent biophysical environments and features.[58-60] Considering the above determinants, environmental education and awareness campaigns, advocacy, and communication are vital to support conservation interest and willingness to invest. Stakeholder relations and effective community engagements stimulate a willingness to join conservation initiatives. The establishment of economic-based instruments that offer incentives is therefore instrumental in stimulating willingness to invest in EI. However, these must be intertwined with regulatory and law enforcement approaches that aim to protect the EI.

Institutional support mechanisms and policy regime

The results revealed that diverse investment instruments are executed by either public or private institutions to manage EI in both public and private landscapes. Major institutional investments are financial mechanisms such as incentives and policy frameworks for improved governance. Through these investments, conservation programmes and funds are established to achieve conservation goals for EI while enhancing economic and social prosperity.[61] This dual approach is essential amongst developing economies, particularly in South Africa where conservation programmes are anticipated to generate both ecological and socio-economic deliverables on one budget.[62] Where 'win-win' solutions are expected, a comprehensive approach is applied to gather different stakeholders to form PPPs to mobilise funds from multiple sources.[41] The popularity of PPPs in the global conservation community has grown due to shrinking funding[39,48] and a desire to strengthen partnerships with the private sector. This investment approach could lead to a long-term collaboration and partnership between different sectors; however, prior to formal agreement, all partners must have an equal understanding of partnership goals and anticipated outcomes to avoid conflict.[39] The key potential role players in these partnerships are envisaged to be communities, government agencies, private landowners or users, private investors, business sector, NGOs and individual civil society members. For the South African context, NRM should reinvigorate institutional relations with other national departments that hold a coinciding conservation mandate. The national departments with a conservation mandate could commit funds and cooperate in the coordination of institutional arrangements that seek to implement and monitor EI restoration and maintenance programmes.

This review showed a shift away from command and control measures which emphasise policy compliance[63] to economic-based instruments which reward voluntary contributors to ES maintenance[22,64,65] (Figure 3). These instruments go beyond compensating private landowners for avoiding practices that potentially damage EI to demonstrable ES delivery. This approach suggests that investments made in EI restoration and maintenance must be justified; therefore NRM programmes should emphasise rewarding conservation interventions based on ecological outcomes.[45,66] Results-based programmes are more justifiable than action-based programmes to provide evidence towards making the ecological investment case.

The South African NRM programmes should define monitoring measures that demonstrate positive ecological outcomes emanating from the interventions. Anticipated outcomes and monitoring protocols must be clarified accordingly. Dedicated data collectors and analysts should define indicators and present the information to key NRM stakeholders and beyond to secure buy-in.[29]

Recommendations and policy implications

Recommendations were extracted from the review to inform conservation policy- and decision-makers about available evidence and insights necessary for 'unlocking and securing EI investments'. These recommendations may vary among countries depending on policies and approaches. In the South African context:

- Ecological infrastructure investment funds should be clear on anticipated outcomes and set a precise monitoring system and indicators. Application of remote sensing satellite imagery coupled with site inspections could be used to monitor ecological changes.[67]
- Biodiversity stewardship programmes are vital in the management of EI. Through a holistic landscape conservation approach, a group of landowners could be assigned an ecological landscape to maintain collectively. Benefits and incentives could be distributed and shared evenly.
- Advocacy and communications should be strengthened to improve awareness, build capacity and awaken conservation interest. Clear messaging on EI rehabilitation benefits could leverage political and social support for EI investment.
- Formation of PPP could scale up investments. Mobilised funds from environmental and water sectors, insurance companies, international funding agencies for climate change, carbon tax funds, corporate institutions, and philanthropists could be used to establish risk mitigation support mechanisms to protect businesses, deliver ES and enhance long-term protection of EI.
- South African national departments with overlapping conservation mandates should break silos by strengthening conservation engagements, coordination of investments and collaborative partnerships.[68]
- Community of practice platforms (e.g. MaReP Forums) should be utilised where knowledge and learning exchange occurs between conservationists, researchers, managers, planners, the private sector, investors and funders. Socio-ecological systems and factors should be analysed, coupled with expert opinions to understand the decision affecting/influencing the keen investors and unenthusiastic parties. Possible investors compared to those who do not take EI into account, can be predicted.

Conclusion

Demanding development needs drive both governments and private landowners to invest and improve EI functionality. The drivers for EI investments are orientated towards improving ES and human well-being. Private landowner contribution is crucial in EI management. Their willingness to engage is stimulated by economic and ecological returns. Financial incentives, compensation, and favourable social conditions encourage private landowners to voluntarily implement interventions and programmes. This review supports and contributes to the evidence-based policy advice by highlighting the measures to instil collaborative partnerships and collective efforts between government and private landowners to maximise investments and expand ecosystem management capacity. Lessons learnt and recommendations made will help policymakers and conservation managers understand effective institutional support mechanisms that have claimed success in 'unlocking and securing EI investments'. More empirical work dedicated to designing economic-based incentives and finance mechanisms is required as well as the demonstration of successes and returns on investments to make the case for investments.

Acknowledgements

We acknowledge funding from the Natural Resources Management programmes of the South African Department of Environmental Affairs and South African National Biodiversity Institute. We also thank the anonymous reviewers for their comments. The electronic databases used for this study were accessed through Stellenbosch University online library. Neither the University nor the authors are responsible for any misuse of the data or assumptions quoted in this article.

Competing interests

We have no competing interests to declare.

Authors' contributions

M.S.M.: Conceptualisation, methodology, investigations, data collection and analysis, writing draft manuscript. C.M.: Conceptualisation, methodology, supervision, review and editing, funding acquisition. T.E.K.: Conceptualisation, methodology, supervision, review and editing. K.J.E.: Conceptualisation, methodology, supervision, review and editing.

References

1. Gallo JA, Pasquini L, Reyers B, Cowling RM. The role of private conservation areas in biodiversity representation and target achievement within the Little Karoo region, South Africa. Biol Conserv. 2009;142(2):446–454. https://doi.org/10.1016/j.biocon.2008.10.025
2. West S, Cairns R, Schultz L. What constitutes a successful biodiversity corridor? A Q-study in the Cape Floristic Region, South Africa. Biol Conserv. 2016;198:183–192. https://doi.org/10.1016/j.biocon.2016.04.019
3. Cumming TL, Shackleton RT, Förster J, Dini J, Khan A, Gumula M, et al. Achieving the national development agenda and the Sustainable Development Goals (SDGs) through investment in ecological infrastructure: A case study of South Africa. Ecosyst Serv. 2017;27:253–260. https://doi.org/10.1016/j.ecoser.2017.05.005
4. Lee JA, Chon J, Ahn C. Planning landscape corridors in ecological infrastructure using least-cost path methods based on the value of ecosystem services. Sustainability. 2014;7564–7585. https://doi.org/10.3390/su6117564
5. Bishop J, Kapila S, Hicks F, Mitchell P, Vorhies F. New business models for biodiversity conservation. J Sustain For. 2009;28(3–5):285–303. https://doi.org/10.1080/10549810902791481
6. Maze K, Barnett M, Botts EA, Stephens A, Freedman M, Guenther L. Making the case for biodiversity in South Africa: Re-framing biodiversity communications. Bothalia Afr Biodivers Conserv. 2016;46:1–8. https://doi.org/10.4102/abc.v46i1.2039
7. Le Maitre DC, Van Wilgen BW, Gelderblom CM, Bailey C, Chapman RA, Nel JA. Invasive alien trees and water resources in South Africa: Case studies of the costs and benefits of management. For Ecol Manage. 2002;160:143–159. https://doi.org/10.1016/S0378-1127(01)00474-1
8. Marais C, Le Maitre D, Frost P. The Working on Fire Programme: Mainstreaming integrated veld and forest fire management into economic development. Paper presented at: XIV World Forestry Congress; 2015 September 7–11; Durban, South Africa.
9. Giordano T, Blignaut J, Marais C. Natural resource management - an employment catalyst: The case of South Africa. Johannesburg: Development Bank of Southern Africa; 2012.
10. Urgenson LS, Prozesky HE, Esler KJ. Stakeholder perceptions of an ecosystem services approach to clearing invasive alien plants on private land. Ecol Soc. 2013;18(1), Art. #26. https://doi.org/10.5751/ES-05259-180126
11. Turpie JK, Marais C, Blignaut JN. The working for water programme: Evolution of a payments for ecosystem services mechanism that addresses both poverty and ecosystem service delivery in South Africa. Ecol Econ. 2007;65:788–798. https://doi.org/10.1016/j.ecolecon.2007.12.024
12. Wilson A, Vickery J, Pendlebury C. Agri-environment schemes as a tool for reversing declining populations of grassland waders: Mixed benefits from Environmentally Sensitive Areas in England. Biol Conserv. 2007;136(1):128–135. https://doi.org/10.1016/j.biocon.2006.11.010
13. Dal Ferro N, Cocco E, Lazzaro B, Berti A, Morari F. Assessing the role of agri-environmental measures to enhance the environment in the Veneto Region, Italy, with a model-based approach. Agric Ecosyst Environ. 2016;232:312–325. https://doi.org/10.1016/j.agee.2016.08.010
14. Kleijn D, Sutherland WJ. How effective are European agri-environment schemes in conserving and promoting biodiversity? J Appl Ecol. 2003;40:947–969. https://doi.org/10.1111/j.1365-2664.2003.00868.x

15. Batáry P, Dicks LV, Kleijn D, Sutherland WJ. The role of agri-environment schemes in conservation and environmental management. Conserv Biol. 2015;29(4):1006–1016. https://doi.org/10.1111/cobi.12536

16. Kok MTJ, Alkemade R, Bakkenes M, Van Eerdt M, Janse J, Mandryk M, et al. Pathways for agriculture and forestry to contribute to terrestrial biodiversity conservation: A global scenario-study. Biol Conserv. 2018;221:137–150. https://doi.org/10.1016/j.biocon.2018.03.003

17. Van Wilgen BW, Richardson DM, Le Maitre DC, Marais C, Magadlela D. The economic consequences of alien plant invasions: Examples of impacts and approaches to sustainable management in South Africa. Environ Dev Sustain. 2001;3:145–168. https://doi.org/10.1023/A:1011668417953

18. Marais C, Wannenburgh AM. Restoration of water resources (natural capital) through the clearing of invasive alien plants from riparian areas in South Africa - Costs and water benefits. S Afr J Bot. 2008;74(3):526–537. https://doi.org/10.1016/j.sajb.2008.01.175

19. Pullin AS, Stewart GB. Guidelines for systematic review in conservation and environmental management. Conserv Soc. 2007;20(6):1647–1656. https://doi.org/10.1111/j.1523-1739.2006.00485.x

20. Arponen A, Heikkinen RK, Paloniemi R, Pöyry J, Similä J, Kuussaari M. Improving conservation planning for semi-natural grasslands: Integrating connectivity into agri-environment schemes. Biol Conserv. 2013;160:234–241. https://doi.org/10.1016/j.biocon.2013.01.018

21. Bennett G, Cassin J, Carroll N. Natural infrastructure investment and implications for the nexus: A global overview. Ecosyst Serv. 2016;17:293–297. https://doi.org/10.1016/j.ecoser.2015.05.006

22. Mills J, Gaskell P, Ingram J, Chaplin S. Understanding farmers' motivations for providing unsubsidised environmental benefits. Land Use Policy. 2018;76:697–707. https://doi.org/10.1016/j.landusepol.2018.02.053

23. Russi D, Margue H, Oppermann R, Keenleyside C. Result-based agri-environment measures: Market-based instruments, incentives or rewards? The case of Baden-Württemberg. Land Use Policy. 2016;54:69–77. https://doi.org/10.1016/j.landusepol.2016.01.012

24. Weikard H, Kis A, Ungvári G. A simple compensation mechanism for flood protection services on farmland. Land Use Policy. 2017;65:128–134. https://doi.org/10.1016/j.landusepol.2017.04.006

25. Feng D, Liang L, Wu W, Li C, Wang L, Li L, et al. Factors influencing willingness to accept in the paddy land-to-dry land program based on contingent value method. J Clean Prod. 2018;183:392–402. https://doi.org/10.1016/j.jclepro.2018.02.142

26. Harrington E, Hsu D. Roles for government and other sectors in the governance of green infrastructure in the U.S. Environ Sci Policy. 2018;88:104–115. https://doi.org/10.1016/j.envsci.2018.06.003

27. Mcwilliam W, Balzarova M. The role of dairy company policies in support of farm green infrastructure in the absence of government stewardship payments. Land Use Policy. 2017;68:671–680. https://doi.org/10.1016/j.landusepol.2017.08.030

28. Asantewaa E, Xavier F. Value orientation and payment for ecosystem services: Perceived detrimental consequences lead to willingness-to-pay for ecosystem services. J Environ Manage. 2018;206:458–471. https://doi.org/10.1016/j.jenvman.2017.10.059

29. Bremer LL, Auerbach DA, Goldstein JH, Vogl AL, Shemie D, Kroeger T, et al. One size does not fit all: Natural infrastructure investments within the Latin American Water Funds Partnership. Ecosyst Serv. 2016;17:217–236. https://doi.org/10.1016/j.ecoser.2015.12.006

30. Galbraith SM, Hall TE, Tavárez HS, Kooistra CM, Ordoñez JC, Bosque-Pérez NA. Local ecological knowledge reveals effects of policy-driven land use and cover change on beekeepers in Costa Rica. Land Use Policy. 2017;69:112–122. https://doi.org/10.1016/j.landusepol.2017.08.032

31. Sheremet O, Ruokamo E, Juutinen A, Svento R, Hanley N. Incentivising participation and spatial coordination in payment for ecosystem service schemes: Forest Disease Control Programs in Finland. Ecol Econ. 2018;152:260–272. https://doi.org/10.1016/j.ecolecon.2018.06.004

32. Westerink J, Jongeneel R, Polman N, Prager K, Franks J, Dupraz P, et al. Collaborative governance arrangements to deliver spatially coordinated agri-environmental management. Land Use Policy. 2017;69:176–192. https://doi.org/10.1016/j.landusepol.2017.09.002

33. Nielsen ASE, Jacobsen JB, Strange N. Landowner participation in forest conservation programs: A revealed approach using register, spatial and contract data. J For Econ. 2018;30:1–12. https://doi.org/10.1016/j.jfe.2017.10.003

34. Schroeder LA, Isselstein J, Chaplin S, Peel S. Agri-environment schemes: Farmers' acceptance and perception of potential "Payment by Results" in grassland - A case study in England. Land Use Policy. 2013;32:134–144. https://doi.org/10.1016/j.landusepol.2012.10.009

35. Latacz-Lohmann U, Hodge I. European agri-environmental policy for the 21st century. Aust J Agric Resour Econ. 2003;47(1):123–139. https://doi.org/10.1111/1467-8489.00206

36. Kleijn D, Baquero RA, Clough Y, Díaz M, Esteban J De, Fernández F, et al. Mixed biodiversity benefits of agri-environment schemes in five European countries. Ecol Lett. 2006;9(3):243–254. https://doi.org/10.1111/j.1461-0248.2005.00869.x

37. Riley M, Sangster H, Smith H, Chiverrell R, Boyle J. Will farmers work together for conservation? The potential limits of farmers' cooperation in agri-environment measures. Land Use Policy. 2018;70:635–646. https://doi.org/10.1016/j.landusepol.2017.10.049

38. Herzon I, Birge T, Allen B, Povellato A, Vanni F, Hart K, et al. Time to look for evidence: Results-based approach to biodiversity conservation on farmland in Europe. Land Use Policy. 2018;71:347–354. https://doi.org/10.1016/j.landusepol.2017.12.011

39. Glumac B, Han Q, Schaefer W, Van Der Krabben E. Negotiation issues in forming public-private partnerships for brownfield redevelopment: Applying a game theoretical experiment. Land Use Policy. 2015;47:66–77. https://doi.org/10.1016/j.landusepol.2015.03.018

40. Ferraz JC, Coutinho L. Investment policies, development finance and economic transformation: Lessons from BNDES. Struct Chang Econ Dyn. 2017;48:86–102. https://doi.org/10.1016/j.strueco.2017.11.008

41. Zhang L, Hu J, Li Y, Shrestha N. Public-private partnership in enhancing farmers' adaptation to drought: Insights from the Lujiang Flatland in the Nu River (Upper Salween) valley, China. Land Use Policy. 2018;71:138–145. https://doi.org/10.1016/j.landusepol.2017.11.034

42. Franks JR, Emery SB. Incentivising collaborative conservation: Lessons from existing environmental Stewardship Scheme options. Land Use Policy. 2013;30(1):847–862. https://doi.org/10.1016/j.landusepol.2012.06.005

43. Welsh R, Webb ME, Langen TA. Factors affecting landowner enrollment in wetland restoration in northeastern New York State. Land Use Policy. 2018;76:679–685. https://doi.org/10.1016/j.landusepol.2018.02.051

44. Hardy MJ, Bekessy SA, Fitzsimons JA, Mata L, Cook C, Nankivell A, et al. Protecting nature on private land using revolving funds: Assessing property suitability. Biol Conserv. 2018;220:84–93. https://doi.org/10.1016/j.biocon.2018.01.026

45. Yang W, Bryan BA, MacDonald DH, Ward JR, Wells G, Crossman ND, et al. A conservation industry for sustaining natural capital and ecosystem services in agricultural landscapes. Ecol Econ. 2010;69(4):680–689. https://doi.org/10.1016/j.ecolecon.2009.11.028

46. Zammit C. Landowners and conservation markets: Social benefits from two Australian government programs. Land Use Policy. 2013;31:11–16. https://doi.org/10.1016/j.landusepol.2012.01.011

47. Froger G, Ménard S, Méral P. Towards a comparative and critical analysis of biodiversity banks. Ecosyst Serv. 2015;15:152–161. https://doi.org/10.1016/j.ecoser.2014.11.018

48. Guillet F, Semal L. Policy flaws of biodiversity offsetting as a conservation strategy. Biol Conserv. 2018;221:86–90. https://doi.org/10.1016/j.biocon.2018.03.001

49. Coralie C, Guillaume O, Claude N. Tracking the origins and development of biodiversity offsetting in academic research and its implications for conservation: A review. Biol Conserv. 2015;192:492–503. https://doi.org/10.1016/j.biocon.2015.08.036

50. Lapeyre R, Froger G, Hrabanski M. Biodiversity offsets as market-based instruments for ecosystem services? From discourses to practices. Ecosyst Serv. 2015;15:125–133. https://doi.org/10.1016/j.ecoser.2014.10.010

51. Lamarque P, Lambin EF. The effectiveness of marked-based instruments to foster the conservation of extensive land use: The case of geographical indications in the French Alps. Land Use Policy. 2015;42:706–717. https://doi.org/10.1016/j.landusepol.2014.10.009

52. Vatn A. Environmental governance - From public to private? Ecol Econ. 2018;148:170–177. https://doi.org/10.1016/j.ecolecon.2018.01.010

53. Forsyth GG, Kruger FJ, Le Maitre DC. National veldfire risk assessment: Analysis of exposure of social, economic and environmental assets to veldfire hazards in South Africa. CSIR Report No: CSIR/NRE/ECO/ER/2010/0023/C. Stellenbosch: CSIR; 2010. Available from: https://www.westerncape.gov.za/assets/departments/local-government/Fire_Brigade_Services/For_the_fire_service/veldfire_risk_report_v11_0.pdf

54. Hansson A, Pedersen E, Weisner SEB. Landowners' incentives for constructing wetlands in an agricultural area in south Sweden. J Environ Manage. 2012;113:271–278. https://doi.org/10.1016/j.jenvman.2012.09.008

55. Nalau J, Becken S, Mackey B. Ecosystem-based adaptation: A review of the constraints. Environ Sci Policy. 2018;89:357–364. https://doi.org/10.1016/j.envsci.2018.08.014

56. Millennium Ecosystem Assessment. Ecosystems and human well-being: Synthesis. Washington, DC: World Resources Institute; 2005.

57. Kanchanaroek Y, Aslam U. Policy schemes for the transition to sustainable agriculture - Farmer preferences and spatial heterogeneity in northern Thailand. Land Use Policy. 2018;78:227–235. https://doi.org/10.1016/j.landusepol.2018.05.026

58. Yeboah FK, Lupi F, Kaplowitz MD. Agricultural landowners' willingness to participate in a filter strip program for watershed protection. Land Use Policy. 2015;49:75–85. https://doi.org/10.1016/j.landusepol.2015.07.016

59. Zhang X, Han L. Which factors affect farmers' willingness for rural community remediation? A tale of three rural villages in China. Land Use Policy. 2018;74:195–203. https://doi.org/10.1016/j.landusepol.2017.08.014

60. Sweikert LA, Gigliotti LM. Evaluating the role of Farm Bill conservation program participation in conserving America' s grasslands. Land Use Policy. 2019;81:392–399. https://doi.org/10.1016/j.landusepol.2018.10.023

61. Rode J, Wittmer H, Emerton L, Schröter-Schlaack C. 'Ecosystem service opportunities': A practice-oriented framework for identifying economic instruments to enhance biodiversity and human livelihoods. J Nat Conserv. 2016;33:35–47. https://doi.org/10.1016/j.jnc.2016.07.001

62. Hejnowicz AP, Raffaelli DG, Rudd MA, White PCL. Evaluating the outcomes of payments for ecosystem services programmes using a capital asset framework. Ecosyst Serv. 2014;9:83–97. https://doi.org/10.1016/j.ecoser.2014.05.001

63. Davis D, Gartside DF. Challenges for economic policy in sustainable management of marine natural resources. Ecol Econ. 2001;36(2):223–236. https://doi.org/10.1016/S0921-8009(00)00251-2

64. Hrabanski M. The biodiversity offsets as market-based instruments in global governance: Origins, success and controversies. Ecosyst Serv. 2015;15:143–151. https://doi.org/10.1016/j.ecoser.2014.12.010

65. Boisvert V. Conservation banking mechanisms and the economization of nature: An institutional analysis. Ecosyst Serv. 2015;15:134–142. https://doi.org/10.1016/j.ecoser.2015.02.004

66. Hajkowicz S. The evolution of Australia's natural resource management programs: Towards improved targeting and evaluation of investments. Land Use Policy. 2009;26(2):471–478. https://doi.org/10.1016/j.landusepol.2008.06.004

67. Pagiola S, Ramírez E, Gobbi J, De Haan C, Ibrahim M, Murgueitio E, et al. Paying for the environmental services of silvopastoral practices in Nicaragua. Ecol Econ. 2007;4:374–385. https://doi.org/10.1016/j.ecolecon.2007.04.014

68. Angelstam P, Barnes G, Elbakidze M, Marais C, Marsh A, Polonsky S, et al. Collaborative learning to unlock investments for functional ecological infrastructure: Bridging barriers in social-ecological systems in South Africa. Ecosyst Serv. 2017;27:291–304. https://doi.org/10.1016/j.ecoser.2017.04.012

AUTHORS:
Grant Carstens[1]
Richard Hay[1]
Michael van der Laan[1]

AFFILIATION:
[1]Department of Plant and Soil Sciences, University of Pretoria, Pretoria, South Africa

CORRESPONDENCE TO:
Grant Carstens

EMAIL:
grant97carstens@gmail.com

DATES:
Received: 02 Aug. 2020
Accepted: 23 Mar. 2021
Published: 29 Sep. 2021

HOW TO CITE:
Carstens G, Hay R, Van der Laan M. Can home gardening significantly reduce food insecurity in South Africa during times of economic distress? S Afr J Sci. 2021;117(9/10), Art. #8730. https://doi.org/10.17159/sajs.2021/8730

ARTICLE INCLUDES:
☒ Peer review
☐ Supplementary material

DATA AVAILABILITY:
☐ Open data set
☐ All data included
☐ On request from author(s)
☐ Not available
☒ Not applicable

EDITORS:
Teresa Coutinho
Salmina Mokgehle

KEYWORDS:
home gardens, malnutrition, food sovereignty, economic crisis

FUNDING:
None

Can home gardening significantly reduce food insecurity in South Africa during times of economic distress?

The novel coronavirus has revealed major impediments in South African food distribution. Existing challenges will be greatly exacerbated by an economic recession projected to be worse than the Great Depression. Home gardens are decidedly utilised to fortify food security and economic resiliency in the face of crises, especially in impoverished communities. For these communities, home garden produce favourably augments diets consisting predominantly of industrially produced staples and the surplus yield can be sold. Despite many campaigns to alleviate food insecurity – some aimed at developing industrial agriculture and others to establish and uplift home gardens – malnutrition and hunger still plague the impoverished. Dissection of these campaigns reveals common flaws in those that failed and key aspects related to those that succeeded, with successful projects even managing to provide a household's total supply of vegetables. One of the crucial failings was a 'top-down' approach that condescended to participants, ignoring existing knowledge, preferences and social consolidation whilst focusing on meticulously consistent packaged methodologies. Successful projects exalted recipients' own bid for food sovereignty and increased individual and community capacity by providing insightful consultation and access to requested necessary inputs. Obstacles especially present in South Africa include drought and collapse of social capital after withdrawal of institutional support. It has been proven possible that these can be overcome with application of technologies, such as rainwater harvesting, and the creation of common cause such as in national drives. This review of the literature clearly reveals that purposefully uplifted home and community gardens alleviate food insecurity.

Significance:

- Citizens aim for food sovereignty in times of economic crises such as will be brought about by the novel coronavirus.
- We assess the potential of the establishment of home and community gardens to alleviate food insecurity in South Africa.
- Home gardens should mainly target the alleviation of malnutrition, producing vegetables to augment cereal-based diets.
- Protection of social capital by institutional networks ensures durability and long-term success of campaigns.
- Rainwater harvesting technology is immensely influential for the success of home gardens in a South African context.

Introduction

Cultivation of food has long been considered the foundation for a civil society's success[1] and citizen participation is an essential ingredient of democratic processes[2]. The last few centuries have been characterised by professional specialisation within populations.[3] The modern farmer, in contrast to the subsistence or small-scale farmer, mass produces one to a few crops, usually staples, solely for income.[4] In South Africa, each specialised commercial farmer feeds, on average, 82 non-farmers of the population, although there exist 2.9 million households involved in smallholder and subsistence agriculture.[5] This demographic is far removed from the greater market.[6] In South Africa, the total number of people suffering from insufficient food totals 12 million or 20%[5], despite the country being a consistent net exporter[7]. Commercial agriculture has led to value chains that are funnelled towards urban centres flowing along socio-economic lines.[8] It has also reduced the need and desire for cultivation of home gardens.[9] Home gardens are defined as plots of cultivated plants maintained typically, but not always, near homes, by individuals or households who have some self-arranged access to land, either through customary or common law.[10] A wide variety of plants are cultivated for many uses.[11] Watkinson and Makdetla[12] reported that only 5% of households in South Africa were still farming for their main source of food and 10% were farming for supplementary food.

On 15 March 2020, a state of national disaster was declared in South Africa due to the COVID-19 pandemic. The country implemented some of the most stringent lockdown measures on the continent in an attempt to contain the spread.[13] These measures have resulted in collective conservative consumption and investment.[14] There has been profound impact on the economic function of the country with the growth forecast being revised downward from 0.7% to 0.4% for 2020, indicating pronounced downturn and recession.[15] The stifled economic activities have fueled rising levels of hunger and desperation in the country[16], with low-wage workers and those in precarious

employment being disproportionately affected. The poorest 60% of South African households now rely more on social grants than on paid employment to obtain food.[17]

The way in which food supply has been organised has long been noted to decrease food sovereignty[18], creating a troubling situation in which resilience to food shortages is strongly determined by financial standing. Reviews reveal that when food is limited, individuals and communities take to reclaiming food sovereignty through home gardening and urban agriculture.[19,20] Food insecurity manifests when food supplies become compromised despite purchase power, social safety welfare nets are ineffective and families cannot produce enough food.[21] Nutrition security should be one of South Africa's main priorities as the country is listed by the World Health Organization (WHO) as a high-burden country with especially high numbers of stunted children.[22] This is attributed to the average household consumption of fruits and vegetables being half the WHO-recommended rate.[23] The majority of the diet of impoverished people in South Africa is seen to be ultra-processed packaged foods known to be sugar-rich, but nutrient poor.[24] This leads to attempts at alleviating starvation that result in increasing incidences of obesity and non-communicable diseases as caloric needs are met but nutritional needs are not.[25] The resulting phenomenon of 'hidden hunger' haunts much of mainstream food aid and marginalisation of home gardens in agricultural policy has often been correlated with an increase in household vulnerability.[26] This is because nutrient-rich fruits and vegetables are replaced by grains that are easily produced through commercial means, processed, and travel well, although they have insufficient levels of many vital nutrients such as Ca, Mg, Zn, folate, vitamin C and several vitamin precursors.[8]

The aim of this review was to urgently evaluate the role that home gardening can play in improving food security in South Africa against the backdrop of poverty and hunger issues that have existed for decades, which are now combined with the global COVID-19 crisis. This pandemic has been predicted to place massive economic strains on livelihoods, with unemployment expected to rise to as much as 50% and the economic climate predicted to drop below that of the Great Depression.[27]

The 'home gardening' concept

Home gardens vary according to format, layout and crop mix.[28] The variability is determined by unique and complex interactions between culture, surrounding ecology, available resources, skills and preferences, and climate variability.[29] The literature largely suggests that the ideal garden structure mostly depends on the socio-economic standing of those who are self-provisioning. In general, richer families cultivate a greater diversity of herbaceous ornamentals and exotic horticultural crops with less economically stable gardeners focusing on staples and crops for uses other than consumption, such as for fibre and medicine.[30]

Many reviews exist for general observation on home gardens and correlation with social indicators rather than empirical testing of specific designs.[19,31-33] This is most likely due to difficulties in collecting data on harvesting. The promise of institutionally supported packaged methodologies is widely acknowledged but largely unquantified. However, home gardening is considered the most significant form of food production for most people in developing countries[34-36] due to a lower intensity of inputs and investment for nutritional return[37]. Although many home gardens have been abandoned due to economic unfeasibility and calorie-focused nutrition, many sources claim home gardens to be an essential augmentation to increase the resilience and livability of impoverished communities.[10,38,39]

Lessons from other countries and time periods

A commonly touted success story of self-provisioning is the 'Victory Garden' and 'War Garden' promoted by the US government during the World Wars. A schematic of one of the many versions of these gardens is shown in Figure 1. The US Department of Agriculture made available a substantial amount of material at the public's request concerning how to successfully cultivate and process various crops for home consumption.

Figure 1: Plan for a 'Victory' or 'War' garden disseminated to the American public in 1914[1] judged to be a successful programme (left), and a schematic disseminated to gardeners in the British 'Dig for Victory' campaign[48] deemed to be over-complicated and less successful (right).

At least 20 million households participated.[20] The widespread undertaking of gardening was largely because it was promoted as a national duty through many corporate channels.[40] The Victory Garden succeeded in maintaining dietary diversity in the face of crisis as they were estimated to have produced 8 million tonnes of produce, amounting to 40% of the domestic demand.[41] Participation declined sharply after both wars, highlighting how community encouragement is key.[20] Successful establishment of Victory Gardens was vigorously supported by a government-driven national campaign that supplied information, and patriotic common cause.

In Cuba, Mexico and Zimbabwe, home gardens are widely used to provide food and generate income in the face of the uncertainties of local monocropping, political volatility and changing ecosystems.[42,43] In Cuba, where the country entered economic crisis due to the collapse of the Soviet Bloc, home gardens rose to meet the demand for economic resiliency as they 'sustained morale' and dietary diversity during ongoing economic crises.[44] It is noted, however, that staples are not included as popularly cultivated plants because the gardens are kept to augment nutrition obtained from ration cards.[44] In Mexico, propagation material and information is largely exchanged through social networks in local populations.[45] These local populations increase the value of crop diversity.[46] In Zimbabwe, due to the arid climate, crop production relies heavily on irrigation.[33] There exists a large subsistence sector out of necessity due to 72.3% of Zimbabweans surviving on less than USD2 per day, although their exact contribution to food production is unknown.[47] Irrigation schemes have been shown to bring huge benefits in the form of stability and enabling year-round production.[33] Drescher et al.[35] noted a difference between the supported gardens and those tended spontaneously. Supported gardens were designed for optimal production, while spontaneous gardens fulfilled multiple functions. Spontaneous designs also contained a far higher diversity of plants with lower production risks.

Evaluating the success of government programmes in Southeast Asia, Martin et al.[31] found that the large diversity of species occurring indigenously in home gardens led to the success of the gardens. Home gardens were an average size of 0.12 ha and were in addition to cash cropping. Those with greater incomes from cash cropping relied less on home gardens for their food staples, although in resource-poor households, an increase in food consumption often correlated with intensification of home gardening.[42]

Lessons from failures in other countries and time periods

In Uganda, home gardens are considered the most primary source of food and are well known to provide dietary diversity for impoverished people.[32] Whitney et al.[32] concluded that the resilience conferred by these home gardens is threatened by government development programmes. This, together with the co-threat of drought, was causing the loss of traditional knowledge and crops resulting in decreased food sovereignty. The ambitious development programmes trade biodiversity for productivity, curbing resiliency yet increasing dependence on centralised food systems.

Just as the 'Victory Gardens' were considered successful, so too was the 'Dig for Victory' campaign in Great Britain. Domestic vegetable production increased from 2 million to 3 million tonnes during the movement.[48] Analysis reveals that the success was mostly due to spontaneous citizen-led action. Only 10% of gardeners are documented to have used the information made available by government and only 34% named the campaign as their motivation.[49] The 'Growmore Leaflets' released are known for being overly-specific and conveying the general fear of the 'muddled agrarianism' that the masses would practise.[50] Campaign leaders wanted to enforce specific production systems, shown in Figure 1, in comparison to the Victory Gardens, that made information available upon public request.[49] What resulted was an alienation of the population.[48] Although home gardening was successfully used for self-provisioning in the face of crisis, failure of the Dig for Victory campaign itself reveals the need for inclusion of participants.

When the East Bwabwata National Park in Namibia was established in 2007, the native Khwe San were no longer permitted to hunt and gather. In compensation, the government promised assistance through the establishment of subsistence agriculture. Despite this assistance, they remain dependent on government food aid[51], with a reliance on grains, which are generally considered nutritionally inferior[52]. The gardening programme was first established with a community approach, yet due to stealing and a lack of cooperation, this quickly dissolved into individual home gardens. Other challenges faced by the gardeners included lack of infrastructure for protection from animals, complicated irrigation because of long distances to water sources and lack of seed.[51] As well as exclusion from aid and market, many home gardeners were unfamiliar with the vegetables provided. This developing case exhibits the importance of continuous communication channels.

Despite home gardens being described as viable livelihood strategies[43,53,54], worrying observations have been made regarding developing dependency syndromes[54,55]. Gardens that were flourishing with direct contact of the NGO rapidly shrank following withdrawal and eventually folded due to lack of new seed as well as social factors, such as community disputes and crop theft.[54] Successful campaigns keep gardeners accountable and remain instructive whilst not ignoring cultural preferences.[55]

Recent and current operations in South Africa

Currently in South Africa, the combined area of home gardens of rural homesteads amounts to 200 000 ha.[37] This number does not consider urban areas, but it is noted that 72% of South Africa's impoverished people live in rural areas. Production fluctuates wildly[26,56] – primarily due to water availability – which severely undermines the capacity to alleviate food insecurity. Crops grown include maize, sorghum (*Sorghum bicolor*), sweet potatoes (*Ipomoea batatas*), cabbage (*Brassica oleracea* var. *capitata*), squashes (*Cucurbita* sp.) and spinach (*Spinacia oleracea*).[38] Figure 2 shows how direct sales and own consumption make up 41% of the produce distribution in South Africa, indicating a significant portion of production exhibits similar distribution channels of home-grown produce and that there already exists a culture, whether inherent or compelled economically, of buying locally and from small-scale production. This demonstrates willingness for people to consume and purchase home-grown produce compared to countries where most citizens consume produce that is highly travelled and has had to meet many regulatory standards. Common reasons for not home gardening are the lack of labour and inputs. Many attempts to implement home gardening have failed and drought is thought of as a major factor for decreasing garden production and causing food prices to rise, as is lack of community cooperation, community feedback and hasty withdrawals by institutions.[57,58]

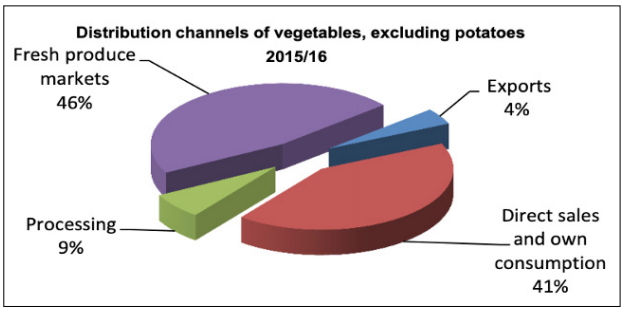


Figure 2: Destinations of vegetables domestically produced in South Africa.

The fate of projects across KwaZulu-Natal is well documented. Modi[56] tested year-round organic production under virtual dryland conditions. Plots of 0.1 ha or 1000 m^2 were assigned to individuals who were given finances for the year amounting to ZAR3000 for seeds and seedlings, and ZAR2400 for labour.[59] Assessments of the harvests revealed that fresh vegetable yields of 19–27 t/ha were achieved with year-round production; variability was attributed to climatic differences influencing

yields because of the dryland approach. Of the total yields from each plot, 12–22 t/ha was consumed by beneficiaries and income amounting to ZAR6000–ZAR15 000 was generated from selling the rest. It was concluded that, under the directive of an informed campaign, crop yields from 0.1 ha would be adequate to meet the food security needs of an average household.[56] Using rainfed systems is risky though, as depicted by Hadebe[60] who evaluated home gardens in the Umkhanyakude district that usually receive similar average rainfall to those sites evaluated by Modi[56] but, at the time of study, experienced a severe drought. Non-irrigators made no income from sales of produce although gardens did satisfy home needs.

Some community garden projects have been considered failures as participants did not see their food security ensured. Shisanya and Hendriks[61] observed that although the gardeners did not exceed a certain threshold of perception of food availability, access to food was nonetheless improved and negative consumption habits avoided. Shisanya and Hendriks[61] also detected flaws in the designs and establishment of these gardens, specifically: plots were too small, participants were given little to no agricultural and nutritional advice, and crop theft was rife. In KwaMashu Township, there is the One Home One Garden programme which started in 2009. By 2010, the programme had established 11 530 household gardens, 30 community gardens and 88 institutional gardens, although evaluation revealed very few households experiencing alleviation of food insecurity and those that did 'governed their gardens independent of state support'.[3] The reason for this is that beneficiaries were completely excluded from decision-making processes.[62] By comparison, in a One House One Farm project implemented in rural Bangladesh which succeeded, participants chose their seeds.[63] Further differences and reasons for failure of the One Home One Garden programme include: unidirectional communication as well as the prevalence of gardener apathy fuelled by state grants.[62]

When evaluating home gardens in the Eastern Cape, Adekunle[21] noted how important they were at providing food on an almost daily basis. Over 60% of rural poor were observed to depend on their own produce year-round. Notably, the majority of successful self-provisioners were married (70.4% and 65% of the male and female participants, respectively).[21] This was attributed to the fact that labour inputs could be delegated. A notable unsuccessful project was the Food Security Community Gardens in Cape Town. Investigations into the failure of these gardens found that whilst they attempted to address lack of skills, land and water, the project neglected the need for leadership and effective communication.[64] It was recommended by Zenda[65] that projects require greater institutional integration.

A particular success story from the Eastern Cape is the Lusikisiki project where the cultivation of β-carotene-containing vegetables was promoted in a combined crop-based and educational intervention.[66] Reviews of the project showed that 'children in the project households experienced alleviation of deficiency symptoms'[66]. Further studies of this dual approach show that virtually all participants eventually obtain substantial food from home gardens with children from project households consuming vegetables more frequently than children from control households.[22] Further, gardening is seen to persist after careful withdrawal focusing on establishing a sustainable source of seed.[66] Another applicable success case is the Siyazondla Homestead Food Production Programme in the Nkonkobe municipality established in 2004/2005. Inputs such as seeds, tools, fencing and water harvesting equipment were disseminated.[67] This programme included urban and rural participants and, although participation varied, vegetable consumption was seen to increase significantly for beneficiaries with 37% reporting that they obtained all their vegetables from their gardens.[68] Notably, 13% of participants reported sharing inputs and produce with non-participants, showing how projects can beneficially spillover.[67] There were, however, a few points for improvement. It was found that only 30.8% of participants used the water harvesting tanks to irrigate.[69] A larger number of tanks, and better demonstration of their effective use, were required. Further, it was found that 85% of participants bought seed over-and-above those provided and cited a shortage as the reason.[67]

In Johannesburg, Gauteng, the Homestead Food Programme began in 1997 as a response to poverty and malnutrition[70] and consisted of the dissemination of production packages containing tools and seed packets[71]. Participants were also given three days' training. The programme reduced food insecurity by 41.5% as gardeners consumed produce directly and generated income from sales.[70] Papers reviewing the project have noted that there was a significant selection bias wherein families that experienced greater food insecurity were more likely to participate.[71] The programme exhibits how food production systems centred around individual households are more reliable and sustainable than standard nutrition interventions such as food aid or state grants.[72]

It has been claimed that the full potential of many home garden programmes is not reached due to lack of exploitation of rainwater harvesting technology.[37,73] Rainwater harvesting is defined as the process of concentrating, collecting and storing rainwater for use at a later time.[74] When evaluating rainwater harvesting use in the Umlazi River catchment, Everson et al.[75] demonstrated that optimal application increased yields by as much as 40–60%, resulting in 75% of farmers finding that community gardens alleviated poverty.

Outlook for South Africa

There are adjunct benefits to home gardening that go beyond food security. It is well known that home gardening can greatly benefit well-being and mental health as well as support plant genetic diversity with well-managed gardens containing as many as 60 different vegetable plant species.[28] Home gardens can mitigate urban heat island effects, regulate stormwater run-off and maintain soil carbon stocks and other aspects of soil quality.[10] They also preserve indigenous knowledge and culture[19] and present opportunities for education and community engagement.

Home gardens are the default in South Africa to supplement food procured with income by those below the poverty line. They buffer vulnerable portions of the population in times of economic downturn.[38] Often, 12–20 people organise themselves into a co-op cultivating 1–2 ha.[60] Many home gardens do not meet their production potential due to factors that are commonly associated with being impoverished: lack of capital, training and education and ability to assume risk. Most programmes aiming at advancing home gardens target specifically rural populations when urban sectors could experience the same benefits, albeit considering site-specific land availability. This is demonstrated by the fact that the average garden size in rural Limpopo is almost double that of the Eastern Cape.[26]

Projecting needs from home gardens

The WHO recommends at least five servings of 80 g, totalling 400 g, of vegetables per person per day.[76] Assuming South African home garden yields of cabbage and Swiss chard of 33 t/ha, beans of 1.4 t/ha, tomatoes of 11 t/ha, and potatoes of 25 t/ha[56,60,69] to reach a grand average of 18 t/ha or 1.8 kg/m^2, and three crops per year to achieve production of 5.4 kg/m^2, each person will need around 27 m^2 of land for their vegetable requirements, or an area roughly 5 × 5 m^2. This may also require around 20 000–25 000 L of water per year depending on growing conditions. Assuming 100 g of vegetables contains approximately 65 calories, a person's vegetable requirement will also satisfy 260 calories for a particular day. As well as accounting for just over 10% of a person's daily energy requirement of 2400 calories per day, the vegetables will satisfy many micronutrient deficiencies, especially Fe and vitamin A, serving as necessary augmentation for a diet in which caloric demand is mostly met by industrially produced cereals. These calculations are complex and highly dependent on site-specific conditions but do have value in moving toward quantitatively determining the potential of home gardens. While an area of over 20 m^2 per person may be unrealistic in many cases, even much smaller areas can contribute to addressing micronutrient deficiencies.

Conclusions

From this review and simple calculation, home gardens show potential for alleviating food insecurity in South Africa in the face of an economic

crisis. Lessons from case studies show clear measures to improve the chance of success. The most important resource is the establishment of networks facilitating two-way dialogue between growers and facilitating institutions to overcome knowledge gaps and deficiencies in capacity.[22] Clearly returning a community's food sovereignty to some degree can bring huge benefit, with the most successful projects proving that it is possible to meet the total demand for vegetables with a home garden.

We therefore conclude that government and/or private institution driven campaigns to promote home gardens can play a significant role in addressing economic and food security challenges. Inciting common cause, whilst avoiding the failings of rigid 'top-down' projects, will be key to successfully transferring packaged methodologies. For doing so in South Africa, we recommend:

- Providing funding for rainwater harvesting systems for home gardens.
- Embracing indigenous knowledge and incorporating crop preferences of local communities.
- Committing to long-term engagement with eventual careful withdrawal accompanied by formation of local committees.
- Evaluation needs to recognise how to target knowledge, planting material and capital dissemination productively.
- Designing proper response to incoming community feedback.
- Gardening being promoted as a national duty to ensure social consolidation and ongoing contact.
- Providing a stable supply of diverse propagation material.
- Provisioning according to gardener inexperience and lack of access to inputs.
- Applying measures of success beyond just yield, such as nutrition profiles, especially in early stages.
- Providing across levels of social differentiation.

Competing interests

We have no competing interests to declare.

Authors' contributions

G.C.: Writing the initial draft, review of literature. R.H.: Conceptualisation, writing revisions. M.v.d.L.: Writing revision, validation, conceptualisation.

References

1. Harrar G, Smith LG, Snyder JC. Victory gardens: Farm. Washington: State College of Washington; 1943.
2. Chihambakwe MT. Deliberative or instrumental participation?: Perceptions of households on the development and implementation of the One Home One Garden programme in KwaMashu Township, [master's thesis]. Durban: University of KwaZulu-Natal; 2014. http://hdl.handle.net/10413/14396
3. South African Department of Agriculture, Forestry and Fisheries (DAFF). Trends in the agricultural sector [document on the Internet]. c2018 [cited 2020 Jul 28. Available from: https://www.dalrrd.gov.za/Portals/0/Statistics%20and%20Economic%20Analysis/Statistical%20Information/Trends%20in%20the%20Agricultural%20Sector%202018.pdf
4. Galor O. From stagnation to growth: Unified growth theory. In: Aghion P, Durlauf SN, editors. Handbook of economic growth. Volume 1 part A. Amsterdam: Elsevier; 2005. p. 171–293. https://doi.org/10.1016/s1574-0684(05)01004-x
5. South African Department of Agriculture, Forestry and Fisheries (DAFF). Annual report 2015/16 [document on Internet]. c2016. [cited 2020 Aug 10]. Available from: https://www.gov.za/sites/default/files/gcis_document/201610/daff-annual-report-2015-2016a.pdf
6. Lahiff E, Cousins B. Smallholder agriculture and land reform in South Africa. IDS Bull. 2005;36(2):127–131. https://doi.org/10.1111/j.1759-5436.2005.tb00209.x
7. Ntombela SM, Bohlmann HR, Kalaba MW. Greening the South Africa's economy could benefit the food sector: Evidence from a carbon tax policy assessment. Environ Resour Econ. 2019;74(2):891–910. https://doi.org/10.1007/s10640-019-00352-9
8. Broadley MR, White PJ. Eats roots and leaves. Can edible horticultural crops address dietary calcium, magnesium and potassium deficiencies? Proc Nutr Soc. 2010;69(4):601–612. https://doi.org/10.1017/s0029665110001588
9. Odebode S. Assessment of home gardening as a potential source of household income in Adinyele Local Government are of Oyo State. Niger J Hortic Sci. 2006;11(1):47–55.
10. Drescher AW, Holmer RJ, Iaquinta DL. Urban homegardens and allotment gardens for sustainable livelihoods: Management strategies and institutional environments. In: Kumar BM, Nair PKR, editors. Tropical homegardens: Advances in agroforestry. Volume 3. Dordrecht: Springer; 2006. p. 317–338. https://doi.org/10.1007/978-1-4020-4948-4_18
11. Peyre A, Guidal A, Wiersum KF, Bongers F. Dynamics of homegarden structure and function in Kerala, India. Agrofor Syst. 2006;66(2):101–115. https://doi.org/10.1007/s10457-005-2919-x
12. Watkinson E, Makgetla N. South Africa's food security crisis. Report for National Labour and Economic Development Institute [document on the Internet]. c2002 [cited 2020 Aug 13]. Available from: https://sarpn.org/documents/d0000077/P93_safscrisis.pdf
13. Renzaho A. The need for the right socio-economic and cultural fit in the COVID-19 response in sub-Saharan Africa: Examining demographic, economic political, health, and socio-cultural differentials in COVID-19 morbidity and mortality. Int J Environ Res Public Health. 2020;17(10):3445. https://doi.org/10.3390/ijerph17103445
14. Ozili PK, Arun T. Spillover of COVID-19: Impact on the global economy [preprint]. SSRN 3562570. 2020. https://doi.org/10.2139/ssrn.3562570
15. Ataguba JE. COVID-19 pandemic, a war to be won: Understanding its economic implications for Africa. Appl Health Econ Health Policy. 2020;18:325–328 https://doi.org/10.1007/s40258-020-00580-x
16. Singh JA. How South Africa's Ministerial Advisory Committee on COVID-19 can be optimised. S Afr Med J. 2020;110(6):439–442. https://doi.org/10.7196/SAMJ.2020.v110i6.14911
17. Baldwin-Ragaven L. Social dimensions of COVID-19 in South Africa: A neglected element of the treatment plan. Wits J Clin Med. 2020;2(SI):33. https://doi.org/10.18772/26180197.2020.v2nsia6
18. Windfuhr M, Jonsén J. Food sovereignty: Towards democracy in localized food systems. Warwickshire: ITDG Publishing; 2005. https://doi.org/10.3362/9781780441160
19. Galhena DH, Freed R, Maredia KM. Home gardens: A promising approach to enhance household food security and wellbeing. Agric Food Secur. 2013;2(1):1–13. https://doi.org/10.1186/2048-7010-2-8
20. Schupp JL, Sharp JS. Exploring the social bases of home gardening. Agric Hum Values. 2012;29(1):93–105. https://doi.org/10.1007/s10460-011-9321-2
21. Adekunle OO. The role of home gardens in household food security in Eastern Cape: A case study of three villages in Nkonkobe Municipality. J Agric Sci. 2013;5(10):67. https://doi.org/10.5539/jas.v5n10p67
22. Faber M, Witten C, Drimie S. Community-based agricultural interventions in the context of food and nutrition security in South Africa. S Afr J Clin Nutr. 2011;24(1):21–30. https://doi.org/10.1080/16070658.2011.11734346
23. Maunder E, Meaker J. The current and potential contribution of home-grown vegetables to diets in South Africa. Water SA. 2007;33(3):401–406. https://doi.org/10.4314/wsa.v33i3.4914–5
24. Corburn J, Vlahov D, Mberu B, Riley L, Caiaffa WT, Rashid SF, et al. Slum health: Arresting COVID-19 and improving well-being in urban informal settlements. J Urban Health. 2020;97:348–357. https://doi.org/10.1007/s11524-020-00438-6
25. Stevano S, Johnston D, Codjoe E. The urban food question in the context of inequality and dietary change: A study of schoolchildren in Accra. J Dev Stud. 2020;56(6):1177–1189. https://doi.org/10.1080/00220388.2019.1632434
26. Shackleton C, Paumgarten F, Cocks M. Household attributes promote diversity of tree holdings in rural areas, South Africa. Agrofor Syst. 2008;72(3):221–230. https://doi.org/10.1007/s10457-007-9066-5

27. Nicola M, Alsafi Z, Sohrabi C, Kerwan A, Al-Jabir A, Iosifidis C, et al. The socio-economic implications of the coronavirus and COVID-19 pandemic: A review. Int J Surg. 2020. https://doi.org/10.1016/j.ijsu.2020.04.018

28. Asfaw Z. Home gardens in Ethiopia: Some observations and generalizations. In: Watson JW, Eyzaguirre, PB, editors. Home gardens and in situ conservation of plant genetic resources in farming systems. Proceedings of the Second International Home Gardens Workshop; 2001 July 17–19; Witzenhausen, Germany. Rome: International Plant Genetic Resources Institute; 2002. p. 125–139. https://doi.org/10.1017/s0014479702251054

29. Egerer MH, Lin BB, Threlfall CG, Kendal D. Temperature variability influences urban garden plant richness and gardener water use behavior, but not planting decisions. Sci Tot Environ. 2019;646:111–120. https://doi.org/10.1016/j.scitotenv.2018.07.270

30. Poot-Pool WS, Van Der Wal H, Flores-Guido S, Pat-Fernández JM, Esparza-Olguín L. Economic stratification differentiates home gardens in the Maya village of Pomuch, México. Econ Bot. 2012;66(3):264–275. https://doi.org/10.1007/s12231-012-9206-3

31. Martin M, Geiger K, Singhakumara BMP, Ashton MS. Quantitatively characterizing the floristics and structure of a traditional homegarden in a village landscape, Sri Lanka. Agrofor Syst. 2019;93(4):1439–1454. https://doi.org/10.1007/s10457-018-0254-2

32. Whitney CW, Luedeling E, Tabuti JR, Nyamukuru A, Hensel O, Gebauer J, et al. Crop diversity in homegardens of southwest Uganda and its importance for rural livelihoods. Agric Hum Health. 2018;35(2):399–424. https://doi.org/10.1007/s10460-017-9835-3

33. Maroyi A. Traditional homegardens and rural livelihoods in Nhema, Zimbabwe: A sustainable agroforestry system. Int J Sustain Dev Worl Ecol. 2009;16(1):1–8. https://doi.org/10.1080/13504500902745895

34. Drimie S. Food security and HIV/AIDS in southern Africa: Case studies and implications for future policy. Report prepared for Action Aid. Pretoria: HSRC; 2004. http://repository.hsrc.ac.za/handle/20.500.11910/7219

35. Drescher A, Hagmann J, Chuma E. Homegardens – a neglected potential for food security and sustainable land management in the communal lands of Zimbabwe. J Agric Rural Dev Trop Subtrop. 1999;100(2):163–180.

36. Nair P. Tropical homegardens: A time-tested example of sustainable agroforestry. Adv Agrofor. 2006. https://doi.org/10.1007/978-1-4020-4948-4

37. Baiyegunhi LJS. Determinants of rainwater harvesting technology (RWHT) adoption for home gardening in Msinga, KwaZulu-Natal, South Africa. Water SA. 2015;41(1):33–39. https://doi.org/10.4314/wsa.v41i1.6

38. Cousins B, Aliber M. Unworkable land reform project designs offer inappropriate farming models to rural dwellers. Cape Town: PLAAS, University of the Western Cape; 2013.

39. Andrew M, Ainslie A, Shackleton C. Land use and livelihoods: Have they been enhanced through land reform? Cape Town: PLAAS, University of the Western Cape; 2003. http://repository.uwc.ac.za/xmlui/bitstream/handle/10566/4245/pb_5_land_use_rural_livelihoods_2003.pdf?sequence=1&isAllowed=y

40. Cole G. Gardening for victory: Victory gardens in American popular periodicals during World War II. The North Dakota Quarterly. 1993;61(3):163–176.

41. Sumner J. Plants go to war: A botanical history of World War II. Jefferson, NC: McFarland; 2019.

42. Danoesastro H, editor The role of homegardens as a source of additional daily income. Paper presented at: Seminar on the Ecology of Homegardens III; 1980 October 25–26; Bandung, Indonesia.

43. Svotwa E, Manyanhaire I, Makombe P. Sustainable gardening on wetlands in the communal lands of Zimbabwe. Electron J Environ Agric Food Chem. 2008;7(3):2754–2760.

44. Buchmann C. Cuban home gardens and their role in social-ecological resilience. Hum Ecol. 2009;37(6):705–721. https://doi.org/10.1007/s10745-009-9283-9

45. Altieri MA, Companioni N, Cañizares K, Murphy C, Rosset P, Bourque M, et al. The greening of the 'barrios': Urban agriculture for food security in Cuba. Agric Hum Values. 1999;16(2):131–140. https://doi.org/10.1023/a:1007545304561

46. Bellon MR. Conceptualizing interventions to support on-farm genetic resource conservation. World Develop. 2004;32(1):159–172. https://doi.org/10.1016/j.worlddev.2003.04.007

47. Davies IA, Torrents A. Overcoming institutional voids in subsistence marketplaces: A Zimbabwean entrepreneurial case. J Macromarketing. 2017;37(3):255–267. https://doi.org/10.1177/0276146717698020

48. Uglow J. A little history of British gardening. London: Pimlico; 2012.

49. Ginn F. Dig for Victory! New histories of wartime gardening in Britain. J Hist Geogr. 2012;38(3):294–305. https://doi.org/10.1016/j.jhg.2012.02.001

50. Spudić S. The new victory garden. Surrey: Royal Horticultural Society; 2007.

51. Mäkelä L. Gardening opportunities as a part of the Khwe San people's food security in the East Bwabwata National Park. Helsinki: University of Helsinki; 2018.

52. Boden G. The Khwe and West Caprivi before Namibian independence: Matters of land, labour, power and alliance. J Namib Studies Hist Polit Cult. 2009;5:27–71.

53. Mudavanhu C, Zinyandu T, Mudavanhu N, Mazorodze S, Chinyanganya T, Manyani A, et al. Smallholder gardening as a sustainable livelihood strategy in Chikwanda communal lands, Gutu, Zimbabwe. Glob Adv Res J Peace Gend Develop Stud. 2012;2(1):1–13.

54. Matsa M, Dzawanda B. Dependency syndrome by communities or insufficient ingestion period by benefactor organizations? The Chirumanzu Caritas Community Gardening Project experience in Zimbabwe. J Geogr Earth Sci. 2014;2(1):127–148.

55. Williamson S. The dependency syndrome: Pesticide use by African smallholders: A report for PAN UK's pesticides poverty and livelihoods project. London: Pesticide Action Network; 2003.

56. Modi AT. A simple model to evaluate integrated vegetable production for food security in KwaZulu-Natal, South Africa. Food Res Int. 2015;76(4):946–952. https://doi.org/10.1016/j.foodres.2015.04.037

57. Quinn CH, Ziervogel G, Taylor A, Takama T, Thomalla F. Coping with multiple stresses in rural South Africa. Ecol Society. 2011;16(3), Art. #2. https://doi.org/10.5751/es-04216-160302

58. Devereux S, Maxwell S. Food security in sub-Saharan Africa. London: ITDG Publishing; 2001. https://doi.org/10.3362/9781780440170.000

59. Nemli S, Asciogul TK, Kaya HB, Kahraman A, Eşiyok D, Tanyolac B. Association mapping for five agronomic traits in the common bean (*Phaseolus vulgaris* L.). J Sci Food Agric. 2014;94(15):3141–3151. https://doi.org/10.1002/jsfa.6664

60. Hadebe N. The impact of capital endowment on smallholder farmers' entrepreneurial drive in taking advantage of small-scale irrigation schemes: Case studies from Makhathini and Ndumo B irrigation schemes in KwaZulu-Natal, South Africa [master's thesis]. Pietermaritzburg: University of KwaZulu-Natal; 2016. https://ukzn-dspace.ukzn.ac.za/bitstream/handle/10413/13975/Hadebe_Nolwazi_2016.pdf?sequence=1&isAllowed=y

61. Shisanya SO, Hendriks SL. The contribution of community gardens to food security in the Maphephetheni. Develop South Afr. 2011;28(4):509–526. https://doi.org/10.1080/0376835x.2011.605568

62. Mthethwa MN. Urban agriculture in Kwamsane, KwaZulu-Natal community and home gardens as an option for food security and poverty reduction [master's thesis]. Durban: University of KwaZulu-Natal; 2012. https://ukzn-dspace.ukzn.ac.za/bitstream/handle/10413/10122/Mthethwa_Menziwokuhle_Ndumiso_2012.pdf?sequence=1&isAllowed=y

63. Ghosh AK, Maharjan KL. Kitchen gardening practices through people's participation: An action research project of three marginalized *bagdi* villages in Bangladesh. J Int Develop Coop. 2013;19(4):107–119. Available from: https://ir.lib.hiroshima-u.ac.jp/files/public/3/35055/20141016204006966O2/JIDC_19-4_107.pdf

64. Mtshisazwe MS. Investigation of the failure of critical food security community gardens as poverty alleviation projects in Cape Town [master's thesis]. Cape Town: Cape Peninsula University of Technology; 2018.

65. Zenda SM. A systems approach to marketing in less developed agriculture with reference to Bululwane Irrigation Scheme [master's thesis]. Alice: University of Fort Hare; 2002. http://hdl.handle.net/10353/d1001045

66. Laurie SM, Faber M. Integrated community-based growth monitoring and vegetable gardens focusing on crops rich in β-carotene: Project evaluation in a rural community in the Eastern Cape, South Africa. J Sci Food Agric. 2008;88(12):2093–2101. https://doi.org/10.1002/jsfa.3319

67. Phezisa B. A situational survey of Siyazondla Homestead Food Production Programme and food security, poverty alleviation in selected communities of Nkonkobe local municipality of the Eastern Cape [master's thesis]. Alice: University of Fort Hare; 2016. http://libdspace.ufh.ac.za/handle/20.500.11837/248

68. Kubheka BP. Impact assessment of the Siyazondla Homestead Food Production Programme in improving household food security of selected households in the Amathole District, Eastern Cape [master's thesis]. Pietermaritzburg: University of KwaZulu-Natal; 2015.

69. Madyibi Z. The Massive Food Production Programme: Does it work? In: Hebinck P, Cousins B, editors. In the shadow of policy: Everyday practices in South Africa's land and agrarian reform. Johannesburg: Wits University Press; 2013. p. 217–230. https://doi.org/10.18772/22013107458.22

70. Rudolph M, Kroll F, Ruysenaar S, Dlamini T. The state of food insecurity in Johannesburg. AFSUN Food Security Series 12. Cape Town: AFSUN; 2016. Available from: http://www.afsun.org/wp-content/uploads/2016/06/afsun12.pdf

71. Tesfamariam BY, Owusu-Sekyere E, Emmanuel D, Elizabeth TB. The impact of the homestead food garden programme on food security in South Africa. Food Secur. 2018;10(1):95–110. https://doi.org/10.1007/s12571-017-0756-1

72. Girard AW, Self JL, McAuliffe C, Olude O. The effects of household food production strategies on the health and nutrition outcomes of women and young children: A systematic review. Paediatr Perinat Epidemiol. 2012;26:205–222. https://doi.org/10.1111/j.1365-3016.2012.01282.x

73. Kahinda JM, Taigbenu A, Boroto RJ. Domestic rainwater harvesting as an adaptation measure to climate change in South Africa. Phys Chem Earth. 2010;35(13–14):742–751. https://doi.org/10.1016/j.pce.2010.07.004

74. Hatibu N, Mahoo HF, Gowing JW, Kajiru G, Lazaro E, Mzirai O. Rainwater harvesting for natural resources management: A planning guide for Tanzania. Nairobi: RELMA; 2000.

75. Everson CS, Everson T, Modi A, Csiwila D, Fanadzo M, Naiken V, et al. Sustainable techniques and practices for water harvesting and conservation and their effective application in resource-poor agricultural production through participatory adaptive research [document on the Internet]. c2011 [cited Aug 02]. Available from: https://www.researchgate.net/profile/Raymond-Auerbach-2/publication/321577153_Sustainable_techniques_and_practices_for_water_harvesting_and_conservation_and_their_effective_application_in_resource-poor_agricultural_production_through_participatory_adaptive_research/links/5a27a36c0f7e9b71dd0cac45/Sustainable-techniques-and-practices-for-water-harvesting-and-conservation-and-their-effective-application-in-resource-poor-agricultural-production-through-participatory-adaptive-research.pdf

76. World Health Organization. Promoting fruit and vegetable consumption around the world. Vienna: WHO Regional office for Europe; 2020. Available from: https://www.euro.who.int/en/health-topics/disease-prevention/nutrition/activities/technical-support-to-member-states/promoting-fruit-and-vegetable-consumption

AUTHORS:
Martins C. Nweke[1]
Adaora J. Okemuo[2]
Ekezie M. Uduonu[2]
Princewill I. Ugwu[3]
Chioma Nwachukwu[3]
Nombeko Mshunqane[1]

AFFILIATIONS:
[1]Department of Physiotherapy, University of Pretoria, Pretoria, South Africa
[2]Department of Medical Rehabilitation, University of Nigeria, Enugu, Nigeria
[3]Department of Physiology, University of Nigeria, Enugu, Nigeria

CORRESPONDENCE TO:
Martins Nweke

EMAIL:
martins.nweke@gmail.com

DATES:
Received: 28 June 2020
Revised: 22 Feb. 2021
Accepted: 16 May 2021
Published: 29 Sep. 2021

HOW TO CITE:
Nweke MC, Okemuo AJ, Uduonu EM, Ugwu PI, Nwachukwu C, Mshunqane N. Meta-analysis of factors affecting prevalence estimates of HIV-associated neurocognitive disorder in sub-Saharan Africa. S Afr J Sci. 2021;117(9/10), Art. #8575. https://doi.org/10.17159/sajs.2021/8575

ARTICLE INCLUDES:
☒ Peer review
☐ Supplementary material

DATA AVAILABILITY:
☐ Open data set
☐ All data included
☐ On request from author(s)
☐ Not available
☒ Not applicable

EDITORS:
Pascal Bessong
Sandiswa Mbewana

KEYWORDS:
HIV infection, neurocognitive disorder, prevalence, sub-Saharan Africa, study characteristics

FUNDING:
University of Pretoria

Meta-analysis of factors affecting prevalence estimates of HIV-associated neurocognitive disorder in sub-Saharan Africa

Successful treatment of HIV with anti-retroviral therapy (ART) is resulting in more people living with HIV-associated neurocognitive disorder (HAND). In sub-Saharan Africa, this calls for strategic planning and judicious allocation of scarce resources, which requires an accurate estimate of the prevalence of HAND. Estimates of the prevalence of HAND in sub-Saharan Africa vary greatly, between 18.8% and 88.3%. This variability may be explained by factors such as different diagnostic approach, neuromedical examination, ART status, sampling method, substance abuse, assessors' qualification, depression and outcome measure. Different methods of diagnosing HAND, different outcome measures and non-random sampling techniques make it almost impossible to accurately estimate the prevalence of HAND in sub-Saharan Africa, often resulting in overestimation of the burden of disease. Consumers of health research should consider certain study characteristics and exercise appropriate caution when interpreting burden of disease in sub-Saharan Africa, especially when pursuing policy shift. Underestimating the prevalence of HAND will certainly affect the capacity and speed of containment, while overestimating will draw unnecessary attention and result in the misallocation of scarce resources.

Significance:

- The high prevalence of HAND in sub-Saharan Africa as estimated in this review calls for further research on the impact of HAND on activities of daily living and putative therapeutic modalities.
- We highlight which study characteristics should be critically checked when using prevalence estimates for the purpose of health policy and distribution of scarce resources in sub-Saharan Africa.
- By favouring certain factors, this review will guide HIV health researchers in which techniques should be used to estimate the burden of HAND. These factors may also apply to estimating the burden of other diseases in sub-Saharan Africa.

Introduction

Human immunodeficiency virus (HIV)-associated neurocognitive disorder (HAND) is one of the most common neurological complications of HIV in the antiretroviral therapy (ART) era.[1] The growing burden of HAND justifies a quintessential global response to address this important mental health challenge among people living with HIV (PLWHIV).[2] Before ART was introduced in 1996, 20–30% of people with advanced HIV infection displayed symptoms of severe HAND[3,4], with death occurring within 6 months[5]. In the ART era, the prevalence of HAND varies between 19% and 64%, averaging globally at about 50%.[6-10] Without a doubt, widespread ART and early intensification with ART are associated with a remarkable drop in HIV-associated dementia.[11] The prevalence of less severe but limiting forms of HAND has continued to increase.[8,12] The rising prevalence of HAND is chiefly associated with HIV becoming a chronic disease because more people are surviving HIV with the use of ART[13], often surviving into their late fifties[14,15]. The chronic nature of HIV has created new challenges for health policymakers and legislators.[16] The neuropsychological complications associated with HIV need to be treated to secure a better well-being and adequate reintegration of PLWHIV into society. The symptoms of HAND include behavioural and cognitive difficulties such as memory loss, poor attention and concentration span, acalculia, poor information processing and inadequate multitasking resulting in poor executive function.[17] Most PLWHIV contract the virus early in life, usually in their late thirties.[18] Long-term treatment and inability to work may have severe economic consequences for many families. Impaired neurocognitive function further predisposes PLWHIV to low productivity, job losses, restricted social participation and poverty.[19]

The recent up-scaling of ART for PLWHIV in sub-Saharan Africa has been appreciable.[20] Consequently, sub-Saharan African healthcare systems need to prepare for and address the surge of chronic complications of HIV such as HAND and accelerated ageing.[20] This will provide a safe trajectory into older adulthood for PLWHIV – a privilege enjoyed by their HIV-seronegative counterparts. More than 50% of the global population of PLWHIV currently resides in sub-Saharan Africa despite comprising less than 10% of the world's population.[21] To ensure a fair quality of life for PLWHIV in sub-Saharan Africa, insight and scholarship are needed for optimal allocation of limited resources. To achieve scholarship in strategic planning for PLWHIV, we need accurate and precise estimates of the prevalence of HAND in sub-Saharan Africa.[22] Underestimating the prevalence of HAND may result in under-budgeting, whereas overestimating may impair the feasibility of such projects by deterring funding agencies. The prevalence of HAND in sub-Saharan Africa has been estimated in several studies[11,23,24]; however, conflicting findings may distort strategic planning.

Hence, there is need for a meta-analysis of the prevalence estimates of HAND in sub-Saharan Africa. Aside from our meta-analysis of HAND prevalence estimates, we explored the effect of different study characteristics such as diagnostic criteria, outcome measures, ART status and duration, assessors' qualifications, assessment

of neuropsychological confounds and other factors on prevalence estimates. Researchers should be able to evaluate the quality of a study based on outcome measure used, the assessor's qualification, sampling technique and other study characteristics.[22] The assessment of outcomes constitutes an important domain in clinical practice, and influences the success of strategic plans and treatment.[25] We reviewed all published estimates of HAND in sub-Saharan Africa.

Materials and methods

Design

This is a systematic review of observation studies including retrospective surveys, cross-sectional studies and cohort studies, focussing on the effects of different outcome measures and assessors on prevalence estimates in sub-Saharan Africa. The protocol was structured using a hybrid of the preferred reporting items for systematic review and meta-analysis (PRISMA) checklist and the *Meta-analysis* of Observational *Studies* in *Epidemiology* (MOOSE) guideline.[26,27] The protocol was registered with Open Science Framework: https://osf.io/vb52y.[28]

Eligibility criteria

Study characteristics

This review included observational studies of epidemiological design written in English, irrespective of sample size and test statistics. Studies had to have been conducted in sub-Saharan Africa.

Participants: This review only included studies in which participants were diagnosed with HAND by a physician or trained person using instruments such as the international HIV dementia scale (IHDS), HIV dementia scale (HDS), Montreal Cognitive Assessment Scale, or neuropsychological battery tests. We included studies irrespective of whether neuropsychological confounds were assessed or whether activities of daily living were evaluated.

Intervention: Not applicable. This is a systematic review of epidemiological studies reporting prevalence of HAND in the sub-Saharan African region.

Control: We included studies irrespective of whether the study had a control group.

Outcomes: For each study, we assessed the estimated prevalence of HAND, sampling method, sub-Saharan African region, measuring instrument and assessors' qualification, neuromedical exam, depression screening, alcohol dependence and substance abuse.

Inclusion criteria

1. Studies conducted among PLWHIV in sub-Saharan Africa reporting prevalence of HAND
2. Studies conducted in which assessment tool and/or assessor's qualification was stated
3. Studies conducted between 2009 and 2019

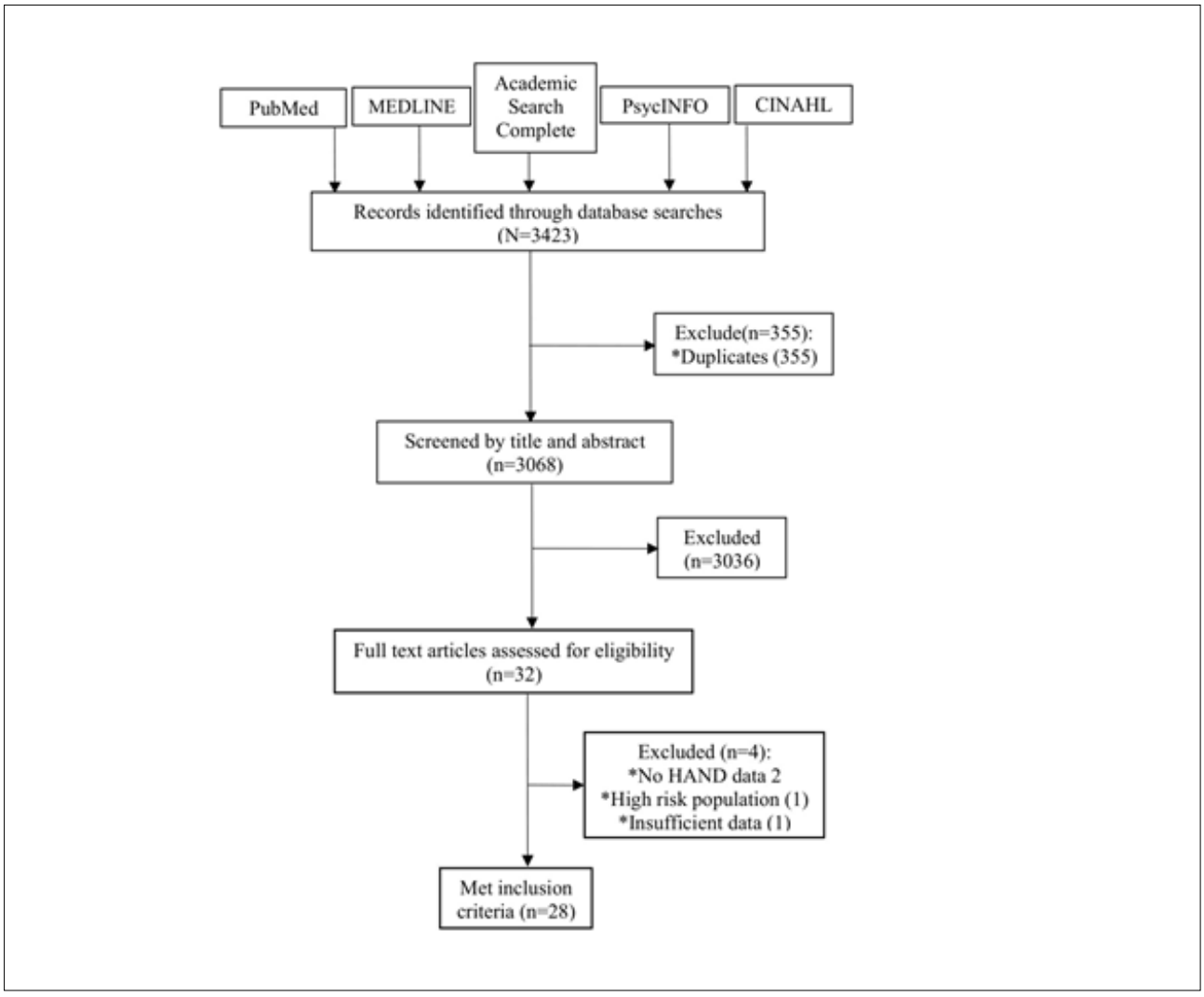


Figure 1: PRISMA flow diagram of the systematic review of articles estimating the prevalence of HIV-associated neurological disorder (HAND) among people living with HIV in sub-Saharan Africa (2009–2019).

Table 1: Studies included in a meta-analysis of the prevalence estimates of HIV-associated neurological disorder among people living with HIV in sub-Saharan Africa (2009–2019). Socio-demographic characteristics of the participants in the studies and study quality.

Study	Sample size	Age (years) Mean±s.d.	Sex (% male)	Education (years)	Study quality
Akolo et al.[55]	133	31.5±7	40.5	13.36 ± 2.40years	Low risk
Anderson et al.[61]	80	35±0.00	20	12±2.00	Low risk
Asiedu et al.[62]	104	37.15±10.06	21.2	11.55±1.55 years	Low risk
Atashili et al.[7]	400	41.0±6	25.8	69% ≤ 6years 39% > 6years	Low risk
Belete et al.[59]	234	18-64	35.5	73.1% ≤ 6years 12.8% >6years 14.1% > 12years	Low risk
Debalkie et al.[44]	684	38.8±8.8	44.0	80.7% ≤ 6years 19.3% > 6years	Low risk
Hestad et al.[63]	275	38.0±8.1	34	9.45±2.3	Low risk
Joska et al.[64]	170	29.5±3.7	25.9	10.06 ± 1.85years	Low risk
Joska et al.[65]	536	34.0±7.9	26.7	45.1% < 9years 54.9% ≥ 10years	Low risk
Joska et al.[36]	55	68.53±4.1	29.7	70% ≤6years 30% >6years	Low risk
Kelly et al.[9]	106	39.0±2.0	27.0	70% ≤ 7years 30% > 7years	Low risk
Lawler et al.[6]	120	37.5±6.5	50.0	37% ≤ 7years 48% 8-12years 15% >12years	Low risk
Milanini et al.[37]	2472	40±11	41	60% ≤6years 40% >6years	
Mogambery et al.[38]	146	35.0±3.3	45.9	NR	Low risk
Mohamed et al.[66]	360	40.2±11.5	35	9.9±3.1	Low risk
Mugendi et al.[45]	345	42.0±9.5	58.6	24.06% ≤ 6 years 47.83% ≤ 12 years 28.11% > 12 years	Low risk
Nakasujja et al.[67]	127	32.2±5.11	33.8	9.7±4,25	Low risk
Nakku et al.[68]	618	NR	27.3	57.3% ≤ 6 years 42.7% > 12 years	Low risk
Namagga et al.[23]	393	37.9±8.7	27	43% <6years 57% >6years	Low risk
Njamnshi et al.[69]	185	37.6±8.8	33	37% ≤ 6years 63% > 6years	Low risk
Nweke et al.[70]	120	NR	24.2	21.67% ≤ 6years 25% ≤12years 53.33% > 12years	Low risk
Nyamayaro et al.[71]	231	37.8±11.2	33.5	19.4% < 7years 80.6% > 7years	Low risk
Patel et al.[40]	179	36.7	35.2	34.6 < 7years 65.4 > 7years	Low risk
Sacktor et al.[39]	77	37.0±3.4	38.0	Average of 8 years	Low risk
Tomita et al.[72]	200	34.5±2.3	19.5	10.88±1.29	Low risk
Tsegaw et al.[73]	593	38.6±10.6	47.9	50.4% < 6years 49.6% > 6years	Low risk
Yakasai et al.[11]	80	36.8±9.0	55.0	Average of 12years	Low risk
Yusuf et al.[74]	418	37.2±9.3	22.2	7.9±6.2years	Low risk

Exclusion criteria

1. Studies conducted among PLWHIV in sub-Saharan Africa but not reporting on prevalence of HAND
2. Studies conducted in which assessment tool and/or assessor's qualification was not stated
3. Studies conducted before 2009

Information sources and search strategy

We searched the literature using several combinations of search terms from medical subject headings (MeSH), and keywords in the title, abstract and/or text of the articles. First, we did a pilot search in PubMed to establish the face sensitivity of the search strategy. The PubMed pilot search included various MeSH terms and keywords/free text terms generated from articles that were key to the research question. We conducted the pilot search using several combinations of the search terms. The most sensitive and specific terms were chosen and reported. The terms were adapted to the syntax and subject headings of the remaining databases. Finally, we searched the following databases: PubMed, MEDLINE, Academic Search Complete, CINAHL and PsycINFO. We also searched the reference lists of identified observational and review articles for relevant studies.

Study records and data management

We exported all literature search results into EndNote 8 and removed all duplicate articles. We screened all bibliographic records in EndNote 8 and then selected articles that met the inclusion criteria. We developed, piloted and refined screening forms that included eligibility questions to aid in the screening process.

Selection process

Initially, one author screened all the articles to identify those that met the inclusion criteria. The rest of the authors then applied the exclusion criteria simultaneously to the database. Any conflicts were then resolved until all authors agreed on the articles to be included. The primary author (MCN) critically cross-checked the initial screening results and read through the full texts of the selected studies for further screening to ensure that the eligibility criteria were met. We emailed the authors of selected studies to clarify issues that impacted on the selection of an article. Details of the flow of studies throughout the selection process, along with the reasons for exclusion, are presented using a PRISMA diagram (Figure 1).

Data collection process

Quality appraisal and risk of bias assessment

We assessed the quality and risk of bias of each article using the quality assessment checklist for prevalence studies adapted from Hoy et al.[29] The checklist examines the appropriateness and adequacy of methodology, study design, participant recruitment, data collection, data analysis and presentation of findings. It is suitable for appraising most studies reporting prevalence. The tool contains 10 items/questions with the 10th item being the summary score. Studies were classified as unclear when there was not enough information to evaluate the risk of bias. All the authors independently assessed the risk of bias, and the primary author (MCN) collated the results.

Data items

Data collected from each article were: authors' affiliations, participants' characteristics, inclusion criteria, exclusion criteria, study sample size, sampling methods, diagnostic criteria/method of assessment, ART status, duration on ART, country, region, instrument used, the assessors' qualification/experience, neuropsychological confounds and results/conclusions.

Data synthesis and assessment of heterogeneity

We used a random-effect model of meta-analysis to estimate the pooled prevalence estimates of HAND, as well as estimates for different measuring instruments and categories of assessors.[30,31] Measures of heterogeneity, i.e. study characteristics, were sorted by year of publication and are presented in an evidence table (Table 1). Measure of heterogeneity, the Cochrane's Q statistics, and I^2 were computed in line with Higgins and Thompson[32]. The I^2-value was interpreted in line with the Cochrane Handbook for Systematic Reviews of Intervention as follows: 0–40% might indicate low heterogeneity, 30–60% may represent moderate heterogeneity, 50–90% may represent substantial heterogeneity, and 75–100% may indicate considerable heterogeneity.[33]

Ethical consideration

This review is related to a clinical trial approved by the Research Ethics Committee of the Faculty of Health Sciences, University of Pretoria (ethics reference number: 152/2020), which complies with the ICH-GCP guidelines and the US Federal wide Assurance.

Data analysis

The pooled prevalence of HAND was estimated using MedCalc. The effects of study characteristics on prevalence estimates of HAND were explored using SPSS version 2, with α set at 0.05. First, we tested the effect of each study characteristic on prevalence using an independent test or ANOVA, and then repeated the test while adjusting for all significant covariates using univariate analysis. To differentiate statistical significance from epidemiological significance, an increase or decrease in prevalence by $\geq$7% was termed epidemiologically significant. A covariate was deemed significant if it accounted for a variation in prevalence of $\geq$7%. We calculated the burden of HAND by multiplying the number of PLWHIV in sub-Saharan Africa by the estimated prevalence of HAND (with 95% confidence intervals [CIs]).[34] We obtained the number of adults with HIV in sub-Saharan Africa (25.6 million) from the 2018 UNAIDS report.[35]

Results

Review profile

We identified 3423 records. After removing duplicates, 3068 records remained. After screening all the titles and abstracts, we excluded 3036 irrelevant records, leaving 32 records for full-text review. Of the 32 full texts, 4 publications were excluded. Ultimately, our review included 28 articles involving 9315 participants from 12 countries (Figure 1). Of these 28 articles, 24 were found in PubMed, 2 in Academic Search Complete, and 1 each from MEDLINE and PsycINFO. The sample sizes in each study varied from 55[36] to 2471[37]. More than half (61%) of the sample population were from East Africa; 39% of the studies were conducted in East Africa, 10 (35.7%) in Southern Africa and 7 (25%) in West Africa. Seventeen studies (71%) used non-probability sampling methods. All studies included adult participants, with a mean age of approximately 38$\pm$7 years. The ratio of women to men was approximately 2:1. Most of the participants had formal education ($\geq$7years of formal education).

Neuro-psychiatric disorder was assessed and excluded in 19 (68%) studies. Comprehensive neuropsychological assessment was carried out in 10 (35.7%) studies, 14 (53.6%) studies used IHDS, 2 (7.1%) employed HDS, and 1 each employed a combination of IHDS and Montreal Cognitive Assessment Scale and a community screening interview for dementia. We noted that different studies used different definitions and criteria to diagnose HAND. Six studies used the 2007 Frascati criteria, with three of them being East African studies. Sixteen (57%) studies diagnosed HAND if participants had $\leq$10 on IHDS/HDS, with most being East African studies. Four (14.3%) studies used the Global Deficit Score. One study from Kenya used an aggregate score derived from the Montreal Cognitive Assessment Scale and IHDS. In terms of methodological quality, risk of bias assessment revealed that all studies had a low risk of bias (Table 1). Of the 28 studies included in this review, 6 reported prevalence of depression, with the prevalence of depression being significantly higher (35.6%) in studies in which depression was associated with HAND than those in which the prevalence of depression (14.3%) was not associated with HAND. In 13 (46.4%) studies, depression was excluded or the rate was not associated with HAND, as opposed to 11 studies in which depression

was included or rate was not associated with HAND. Four studies reported level of alcohol use. Mogambery et al.[38] reported that higher alcohol use (42.5%) was protective of HAND as opposed to studies that reported low levels of alcohol use (13.5%) which found no association between prevalence of HAND and alcohol use. Of the studies that did not outrightly exclude participants with substance abuse, only three reported rate of substance abuse, which was not associated with HAND (Table 1).

Prevalence and burden of HAND in sub-Saharan Africa

The prevalence of HAND in sub-Saharan Africa varies greatly between 14% (9.2 to 19.9%)[39] and 88% (79.0 to 94.5%)[40]. With a considerable degree of heterogeneity, the pooled prevalence was pegged at 53% (CI=44.3 to 61.5%; I^2=98.5). Thirteen (46%) of the studies reported prevalence rates higher than the summary prevalence (Figure 2). The estimated burden of HAND in sub-Saharan Africa was 1356.8 million (m) (CI: 1134.08–1574.4 m). The prevalence of HAND in East Africa, West Africa and Southern Africa was 66.7 m (CI: 35.0–98.4 m), 49.6 m (CI: 23.6–75.6 m) and 48.4 m (CI: 22.5–74.4 m), respectively.

Effects of study characteristics on prevalence of HAND in sub-Saharan Africa

Our data were normally distributed (Shapiro–Wilk statistic=0.940; p=0.111), hence we used parametric statistics. Table 2 shows the effects of study characteristics on the prevalence estimates of HAND in sub-Saharan Africa.

Epidemiologically, studies from East Africa reported significantly higher prevalence estimates of HAND ((66.7 (CI: 35.0–98.4)) compared to studies from Southern Africa and West Africa. Studies that utilised random sampling recorded an epidemiologically significantly lower rate ((46.5% (CI: 26.1–66.9)) compared with studies that used consecutive sampling ((68.4 (CI: 44.8–92.0)). Epidemiologically, similar prevalence estimates of HAND were obtained using comprehensive neuropsychological tests and validated composite measures. Nineteen (68%) studies diagnosed HAND using a neuromedical exam. Prevalence estimates of HAND were significantly lower (44.7% (CI: 32.2–57.3)) in studies with neuromedical exam compared to studies without neuromedical exam (71.3% (CI: 48.0–94.5)). Epidemiologically higher prevalence estimates of HAND were reported in studies in which physicians/PhD holders were the outcome assessors (58.8%) than those studies with paramedics (49.6%). A large proportion of studies did not exclude neuropsychological confounds such as depression (79%), alcohol use (57%) and substance abuse (61%). We classified studies into two categories: (1) studies in which neuropsychological confound(s) were excluded or potential confounds were not associated with HAND prevalence estimates and (2) studies that included neuropsychological confound(s) which were associated with HAND prevalence estimate. Studies that excluded depression and substance abuse reported epidemiologically lower prevalence estimates of HAND (51.9% and 45.7%, respectively) than studies that included depression and substance abuse (p<0.05) (Table 2).

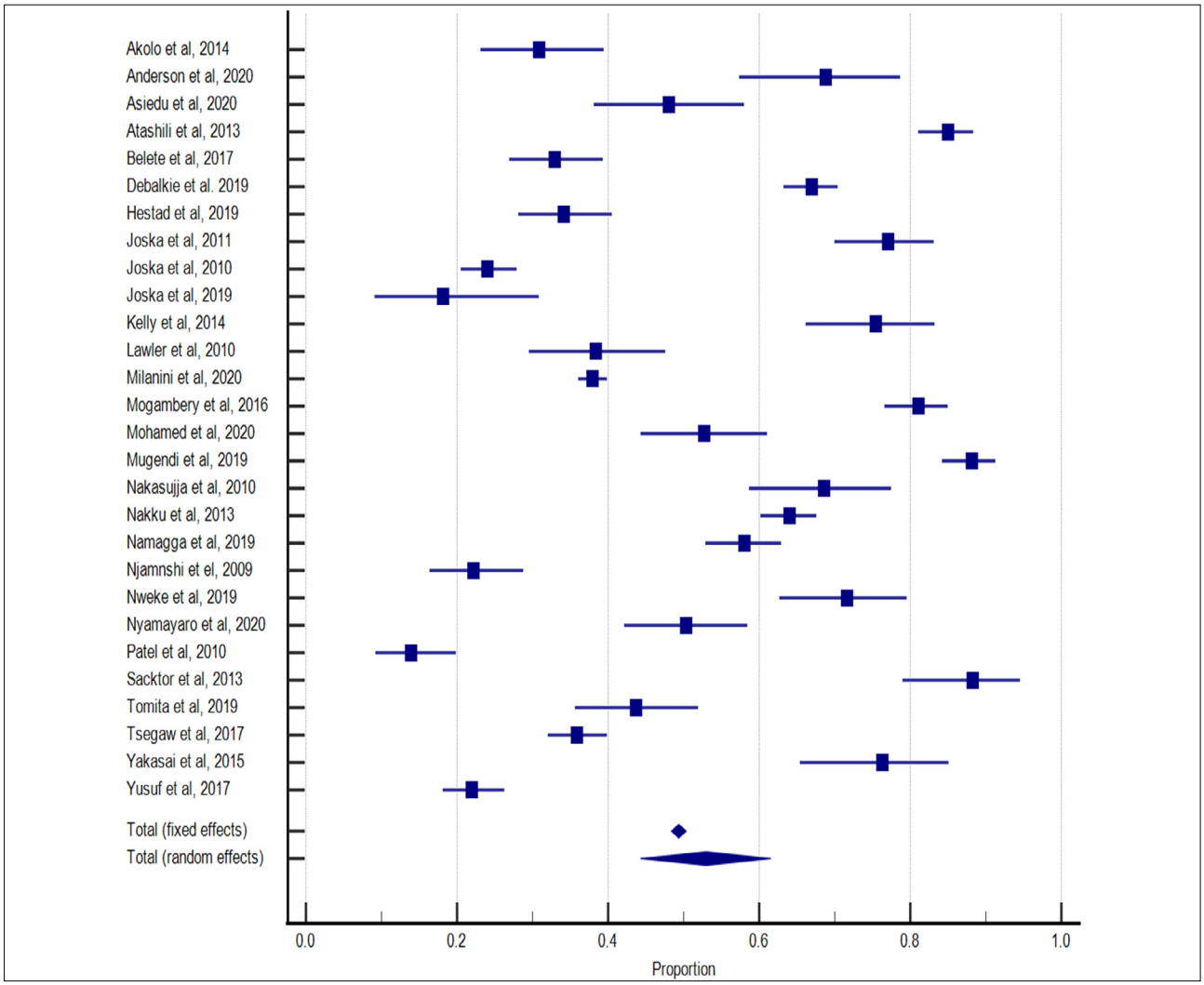


Figure 2: Forest plot displaying pooled prevalence of HIV-associated neurological disorder in sub-Saharan Africa. Estimates are generated from studies published between 2009 and 2019.

Table 2: Effects of study characteristics on prevalence of HIV-associated neurological disorder among people living with HIV in sub-Saharan Africa

Study characteristics	Publications *n* (%)	Prevalence % (95% CI)	t/F-value	P	Adjusted prevalenceα % (95% CI)	F-value	*p*-value
ART status							
Experienced	14 (50)	55.7(42.8–68.7)[a]	1.720	0.200†	64.7 (45.8–83.6)[a]	2.721*†	0.125
Naïve	5 (17.9)	63.5 (35.7–91.7)[b]			41.9 (-0.5–84.3)[b]		
Mixed	9 (32.1)	41.8 (23.9–59.8)[c]			29.4 (0.4–58.3)[c]		
Outcome measure							
Comprehensive NP	10 (35.7)	55.9 (38.9–72.9)	0.307	0.588	58.2 (32.7–83.7)	0.127	0.732
Validated but non-NP	18 (64.3)	50.8 (39.1–62.5)			51.7 (28.0–75.4)		
Diagnostic criteria							
Frascati	6 (21.4)	62.6 (34.9–90.3)[a]	0.751	0.482†	45.4 (12.0–78.8)[a]	4.547*†	0.048
Global deficit score	4 (14.3)	50.8 (39.1–62.5)[b]			69.7 (44.7–94.8)[b]		
IHDS/HDS (≤10)	18 (64.3)	46.0 (18.1–73.8)[bc]			4.6 (-35.8–44.7)[c]		
Sampling technique							
Random	8 (28.6.0)	46.5 (26.1–66.9)[a]	1.254	0.303†	46.5 (19.1–73.9)[a]	2.014†	0.196
Consecutive sampling	7 (25.0)	64.3 (43.4–85.2)[b]			68.4 (44.8–92.0)[b]		
Convenience	10 (46.4)	50.1 (36.6–63.2)[ac]			43.7 (26.6–60.8)[a]		
Neuromedical exam?							
Yes	19 (67.9)	49.6 (38.3–60.9)[a]	1.004	0.326†	44.7 (32.2–57.3)[a]	3.843†	0.074
No	9 (32.1)	59.0 (41.6–76.5)[b]			71.3 (48.0–94.5)[b]		
Outcome assessor							
Paramedics	7 (25.0)	49.9 (35.5–64.4)[a]	0.087	0.917†	39.7 (13.6–65.9)[a]	0.793†	0.477
Physicians/PhD	20 (71.4)	58.8 (41.5–66.1)[b]			56.5 (44.4–68.6)[b]		
Exclusion of neuropsychological confounds							
Depression							
Yes	13 (46.4)	48.1 (38.2–57.9)[a]	2.340	0.117†	51.9 (32.8–71.0)[a]	0.955†	0.415
No	11 (39.3)	63.2 (46.1–80.3)[b]			59.3 (36.2–82.3)[b]		
Missing data	4 (14.3)	38.4 (-9.2–86.0)[c]			37.07 (9.5–64.7)[c]		
Alcoholism							
Yes	17 (60.7)	51.0 (37.7–64.4)[a]	0.897	0.558†	53.0 (38.7–67.2)[a]	0.015	0.985
No	4 (14.3)	64.5 (51.4–77.7)[b]			49.9 (-0.6–100.3)[a]		
Missing data	7 (25.0)	49.8 (29.1–70.4)[ac]			51.0 (30.8–71.2)[a]		
Substance Abuse							
Yes	15 (53.6)	52.3 (39.5–65.0)[a]	0.639	0.536†	45.7 (39.9–143.9)[a]	0.904†	0.434
No	2 (7.1)	70.3 (53.2–87.5)[b]			80.9 (17.8–76.9)[b]		
Missing data	11 (39.3)	49.9 (32.8–67.0)[ac]			58.4 (39.4–76.9)[c]		
Region							
West Africa	7 (25)	50.7 (25.9–76.9)[a]	2.498	0.103†	49.6 (23.6–75.6)[a]	0.525†	0.616
East Africa	11 (39.3)	63.5 (50.0–76.9)[b]			66.7 (35.0–98.4)[b]		
Southern Africa	10 (35.7)	42.0 (27.2–56.9)[c]			48.4 (22.5–74.4)[a]		

**Significant at α=0.05; †: epidemiologically significant at difference of ≥7%.*

Note: Pairs with the same superscripts are not statistically or epidemiologically significantly different, while pairs with different superscripts are statistically or epidemiologically significantly different.

Discussion

The estimated prevalence of HAND in sub-Saharan Africa in this review was approximately 53% (CI: 44.3–61.5%). Unfortunately, none of the reviewed studies assessed for self-report of difficulty in cognitive ability as part of the criteria for diagnosing HAND. It is likely that use of diagnostic criteria void of a patient's perspective of the symptom may have resulted in exaggerated prevalence estimates.[34,41] Given that the false positive rate can exceed 20%.[41,42] We estimate the prevalence of HAND to be pegged at 33%. Our prevalence estimates are 10% higher than estimates obtained in a similar study conducted 7 years ago.[43] This increase reflects the widespread use of ART in sub-Saharan Africa and indicates that asymptomatic and mild cognitive impairment are common amongst PLWHIV.

The prevalence of HAND seems to be higher among ART experienced PLWHIV compared to ART naïve individuals.[14,15] ART naïve PLWHIV showed a higher prevalence of HAND compared to PLWHIV in a mixed ART group, suggesting that duration of use of ART is an important factor influencing the prevalence of HAND and the severity of symptoms. This is consistent with the findings of Habib et al.[43] in which individuals on ART for ≥6 months achieved a lower prevalence of dementia. We included studies from 2009 to 2019 only, and no cases of acute HAND were diagnosed in any of these studies, suggesting that ART has been successful in combating severe forms of neurocognitive impairment among PLWHIV in sub-Saharan Africa. In other words, non-use of ART among ART-naïve PLWHIV is a risk for HAND, while longer duration of ART is a risk for mild forms of HAND among ART-experienced PLWHIV.[14,15] Interestingly, we found an inverse relationship between the prevalence of HAND and the prevalence of HIV. In sub-Saharan Africa, Southern Africa has the highest prevalence of HIV, followed by Eastern Africa and then West Africa. Our review suggests that East Africa had the largest burden of HAND, followed by West Africa and then Southern Africa. This may be due to the South African HIV care system being more effective than those of East and West Africa[44,45] or due to a larger proportion of PLWHIV in South Africa being on ART[46].

Our findings also suggest that the prevalence estimates of HAND were associated with sampling technique. Studies that used consecutive sampling generated significantly higher prevalence estimates compared to studies that employed random sampling. Estimating prevalence using non-probability sampling may be prone to a type-1 error due to selection bias[47], and may have contributed to the wide prevalence gap across studies. We suggest that policymakers cautiously interpret studies that employ non-probability sampling, especially when using such evidence for strategic planning. Random sampling is the gold standard procedure for selecting participants for medical research.[48]

Several measures, ranging from generic to disease-specific measures, are used to assess neurocognitive disorder among PLWHIV. These include the popular mini-mental state exam[48], the HDS[49], community screening interview for dementia[50], IHDS[29], and several combinations of neuropsychological tests such as Hopkins verbal learning test revised, Wechsler adult intelligence scale III symbol search, grooved pegboard, colour trails 2, Wechsler memory scale III spatial span and controlled oral word association test[11]. In this review, cognitive measures included IHDS, HDS, community screening interview for dementia, Montreal Cognitive Assessment Scale and multidomain neuropsychological tests. Consistent with the literature[9,51], we found that studies using comprehensive multidomain neuropsychological battery tests delivered higher prevalence estimates of HAND. Neuropsychological tests are generally more sensitive than composite measures such as IHDS and HDS.[51,52] Generally, HAND is diagnosed with a IHDS/HDS score less than 10. We found that prevalence estimates of HAND were statistically and epidemiologically lower in studies that used the Frascati criteria and Global Deficit Score compared to studies that relied on IHDS/HDS score alone. This suggests that the Frascati criteria are stricter and more specific than the IHDS/HDS score.[39,53] Neuropsychological tests are the gold standard for assessing HAND, especially when the Frascati criteria are fulfilled.[54] The Frascati criteria seems to be gaining wide utility in clinical research across sub-Saharan Africa[40,55]; however, use of composite measures remains the norm amongst clinical researchers in the region[56]. This may be due to battery tests being expensive and health professionals not being skilled in administering these tests to diagnose HAND.[57,58] Our findings suggest that the Frascati criteria may be more sensitive, generating higher prevalence estimates than the Global Deficit Score. Based on this review, we recommend comprehensive neuropsychological screening and application of the Frascati criteria to diagnose HAND for improved sensitivity as well as reduced false positive rate, especially when planning treatment.

We found that studies that conducted neuromedical exams and screened for depression generated significantly lower prevalence estimates. This is consistent with the findings of Belete et al.[59] and Lawler et al.[6] who maintain that neurological and neuropsychiatric confounds, such as depression and a positive history of psychiatric disorder, constitute major bias and often result in exaggerated prevalence estimates. Non-exclusion of neuropsychological confounds such as depression, substance abuse or lack of neurological and/or medical examination may also result in false positive diagnoses. We recommend compulsory neuromedical examination as well as screening for exclusion of putative neuropsychological confounds such as depression, alcoholism and substance abuse. Employing the Frascati criteria may also help to eliminate confounding variables.[4] The Frascati criteria were developed in 2007 to allow for uniformity in the diagnosis of HAND. These criteria involve neuropsychological testing across various cognitive domains. Activities of daily living and ruling out other potential factors/causes for cognitive decline are also part of the diagnostic criteria.[16] Our review reveals a grey area of controversy: whether alcohol use is protective of, or a risk for HAND. Studies by Debalkie Animut et al.[44] and Patel et al.[40] concluded that alcohol use is a significant predictor of HAND prevalence in a population. Interestingly, an important study[38] that found a high rate of alcohol use achieved a reciprocal association between alcohol use and the prevalence of HAND. It is possible that alcohol use in this population negatively influenced ART use (adherence), which might protect against mild forms of HAND.[60] We found that studies with more highly qualified and experienced assessors resulted in higher prevalence estimates of HAND, possibly due to physicians or assessors with a PhD being less likely to miss a diagnosis. This shows that the qualification or experience of outcome assessors is an important consideration when screening for HAND in clinical research across sub-Saharan Africa.

Conclusions

Prevalence of HAND in sub-Saharan Africa is high and widespread, with HAND being most prevalent in East Africa. In sub-Saharan Africa, HAND is estimated under circumstances that differ: using different diagnostic approaches, different degrees of neuromedical examination, across people with different ART status, different sampling methods, assessment of substance abuse, assessors' qualification, depression and different outcome measures. These variables, in this order, account for different prevalence estimates of HAND reported in different studies. We recommend that estimating the prevalence of HAND should be standardised, favouring the use of Frascati criteria and multidomain comprehensive neuropsychological screening, neuromedical examination, factoring in of ART status and duration, random sampling technique, assessing and excluding current history of substance abuse and depression, and inter-professional collaboration with at least one of the assessors being a physician or PhD graduate. Importantly, policymakers should consider study characteristics when interpreting prevalence estimates for strategic planning and policymaking.

Acknowledgements

We acknowledge the Health Promotion, Molecular Therapy and Rehabilitation initiative, University of Nigeria, Enugu, for the motivation to conduct this review. We extend our sincere gratitude to Dr Tosh Cheryl of the Faculty of Health Sciences, University of Ibadan, for editing this manuscript.

Competing interests

We have no competing interests to declare.

Authors' contributions

M.C.N. who is a funded PhD student under the supervision of N.M. conceived the idea and wrote the study protocol. Independent screening was undertaken by A.J.O., E.M.U., and C.N. Differences were resolved in consultation with M.C.N. Data extraction was independently carried out by A.J.O., E.M.U., C.N. and P.I.U., while M.C.N. serves as the data curator. M.C.N. together with A.J.O., C.N., and P.I.U. wrote the initial draft of the manuscript. All the authors revised the initial draft of the manuscript and approved the final draft.

References

1. Sanmarti M, Ibáñez L, Huertas S, Badenes D, Dalmau D, Slevin M. HIV-associated neurocognitive disorders. J Mol Psychiatry. 2014;2(1):371–379. https://doi.org/10.1186/2049-9256-2-2
2. Bougea A, Spantideas N, Galanis P, Gkekas G, Thomaides T. Optimal treatment of HIV-associated neurocognitive disorders: Myths and reality. Ther Adv Infect Dis. 2019;6(8):1–9. https://doi.org/10.1177/2049936119838228
3. Ahmad MY, Mustafa IG. Prevalence and correlates of HIV-associated neurocognitive disorders in Northwestern Nigeria. Neurol Res Int. 2015;2015, Art. #486960. https://doi.org/10.1155/2015/486960
4. Antinori A, Trotta MP, Lorenzini P, Torti C, Gianotti N, Maggiolo F, et al. Virological response to salvage therapy in HIV-infected persons carrying the reverse transcriptase K65R mutation. Antiviral Ther. 2007;12(8):1175–1183.
5. Ances BM, Ellis RJ. Dementia and neurocognitive disorders due to HIV-1 infection. Semin Neurol. 2007;27(1):86–92. https://doi.org/10.1055/s-2006-956759
6. Lawler K, Mosepele M, Ratcliffe S, Seloilive E, Steele K, Nthobatsang R, et al. Neurocognitive impairment among HIV positive individuals in Botswana: A pilot study. J Int AIDS Soc. 2010;13(1):15. https://doi.org/10.1186/1758-2652-13-15
7. Atashili J, Gaynes BN, Pence BW, Tayong G, Kats D, O'Donnell JK. Prevalence, characteristics and correlates of a positive-dementia screen in patients on antiretroviral therapy in Bamenda, Cameroon: A cross-sectional study. BMC Neurol. 2013;13:86. https://doi.org/10.1186/1471-2377-13-86
8. Sacktor N, Skolasky RL, Seaberg E, Munro C, Becker JT, Martin E. Prevalence of HIV-associated neurocognitive disorders in the Multicenter AIDS Cohort Study. Neurology. 2016;86(4):334–340. https://doi.org/10.1212/WNL.0000000000002277
9. Kelly CM, Van Oosterhout JJ, Ngwalo C, Stewart RC, Benjamin L, Robertson KR, et al. HIV associated neurocognitive disorders (HAND) in Malawian adults and effect on adherence to combination anti-retroviral therapy: A cross sectional study. PLoS ONE. 2014;9(6), e98962. https://doi.org/10.1371/journal.pone.0098962
10. Van Wijk C. Screening for HIV-associated neurocognitive disorders (HANDs) in South Africa: A caution against uncritical use of comparative data from other developing countries. S Afr J HIV Med. 2013;14(1):17–19. https://doi.org/10.7196/sajhivmed.855
11. Yakasai AM, Gudaji MJ, Muhammad H, Ibrahim A, Owolabi LF, Ibrahim DA, et al. Prevalence and correlates of HIV-Associated Neurocognitive Disorders (HAND) in Northwestern Nigeria. Neurol Res Int. 2015:1–9. https://doi.org/10.1155/2015/486960
12. Heaton RK, Franklin DR, Ellis RJ, McCutchan JA, Letendre SL, Leblanc S, et al. HIV-associated neurocognitive disorders before and during the era of combination antiretroviral therapy: Differences in rates, nature, and predictors. J Neurovirol; 2011;17(1):3–16. https://doi.org/10.1007/s13365-010-0006-1
13. Deeks SG, Lewin SR, Havlir DV. The end of AIDS: HIV infection as a chronic disease. Lancet. 2013;382(9903):1525–1533. https://doi.org/10.1016/S0140-6736(13)61809-7
14. US Centers for Disease Control and Prevention (CDC). HIV surveillance report vol. 29: Diagnoses of HIV infection in the United States and dependent areas, 2017. Atlanta, GA: CDC; 2018. Available from: http://www.cdc.gov/hiv/library/reports/hiv-surveillance.html
15. UNAIDS. AIDSinfo [webpage on the Internet]. No date [cited 2019 Dec 17]. Available from: https://aidsinfo.unaids.org/
16. Gates TM, Cysique LA, Siefried KJ, Chaganti J, Moffat KJ, Brew BJ. Maraviroc-intensified combined antiretroviral therapy improves cognition in virally suppressed HIV-associated neurocognitive disorder. AIDS. 2016;30(4):591–600. https://doi.org/10.1097/QAD.0000000000000951
17. Modi G, Mochan A, Modi M. Neurological manifestations of HIV. In: Okware SI, editor. Advances in HIV and AIDS control. London: IntechOpen; 2018. https://doi.org/10.5772/intechopen.80054
18. Hall MD. Age transmissions in European HIV transmission pairs uncovered with viral sequence data. Paper presented at: Conference on Retroviruses and Opportunistic Infections (CROI 2018); 2018 March 4–7; Boston, MA, USA. Abstract 960.
19. Alford K, Vera JH. Cognitive impairment in people living with HIV in the ART era: A review. Br Med Bull. 2018;127(1):55–68. https://doi.org/10.1093/bmb/ldy019
20. Wing EJ. The aging population with HIV infection. Trans Am Clin Climatol Assoc. 2017;128:131–144.
21. Nweke MC, Nombeko M, Govender N, Akinpelu AO. Physiologic effects of physical activity on cognitive function in people living with HIV: A systematic review of intervention and observational studies. Afr J Physical Activity Health Sci. 2021;27(1):1–22.
22. Ward MM. Disease prevalence and incidence using administrative data: Some assembly required. J Rheumatol. 2013;40(8):1241–1243. https://doi.org/10.3899/jrheum.130675
23. Namagga JK, Rukundo GZ, Voss JG. Prevalence and risk factors of HIV-associated neurocognitive disorders in Rural Southwestern Uganda. J Assoc Nurses AIDS Care. 2019;30(5):531–538. https://doi.org/10.1097/JNC.0000000000000036
24. Achappa B, Madi D, Bhaskaran U, Ramapuram JT, Rao S. Adherence to antiretroviral therapy in sub-Saharan Africa and North America: A meta-analysis. North Am J Med Sci. 2013;5(3):220–223. https://doi.org/10.4103/1947-2714.109196
25. Kampstra P. Beanplot: A boxplot alternative for visual comparison of distributions. J Stat Softw. 2008;28(1)1–9. https://doi.org/10.18637/jss.v028.c01
26. Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: Elaboration and explanation. BMJ. 2015;349:g7647. https://doi.org/10.1136/bmj.g7647
27. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al. Meta-analysis of observational studies in epidemiology: A proposal for reporting Meta-analysis of Observational Studies in Epidemiology (MOOSE) group. JAMA. 2000;283(15):2008–2012. https://doi.org/10.1001/jama.283.15.2008
28. Nweke MC, Ekezie UM, Chioma N, Adaora O, Ugwu PIMM. Effects of variability in outcome measure and outcome assessors on the prevalence of HAND in sub-Saharan Africa [preregistered protocol]. OSF; 2020. https://doi.org/10.17605/OSF.IO/VB52Y
29. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials. 1986;7(3):177–188. https://doi.org/10.1016/0197-2456(86)90046-2
30. Hoy D, Brooks P, Woolf A, Blyth F, March L, Bain C, et al. Assessing risk of bias in prevalence studies: Modification of an existing tool and evidence of interrater agreement. J Clin Epidemiol. 2012;65(9):934–939. https://doi.org/10.1016/j.jclinepi.2011.11.014
31. Wang KS, Liu X. Statistical methods in the meta-analysis of prevalence of human diseases. J Biostat Epidemiol. 2016;2(1):20–24.
32. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med. 2002;21(11):1539–1558. https://doi.org/10.1002/sim.1186
33. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al., editors. Cochrane handbook for systematic reviews of interventions version 6.2 [document on the Internet]. c2021 [cited 2021 Feb 22]. Available from: https://training.cochrane.org/handbook/current
34. Wang Y, Liu M, Lu Q, Farrell M, Lappin JM, Shi J, et al. Global prevalence and burden of HIV-associated neurocognitive disorder: A meta-analysis. Neurology. 2020;95:e2610–e2621. https://doi.org/10.1212/WNL.0000000000010752

35. UNAIDS. Global HIV AIDS statistics: 2019 fact sheet [webpage on the Internet]. No date [cited 2020 May 08]. Available from: www.unaids.org/en/resources/fact-sheet

36. Joska JA, Dreyer AJ, Nightingale S, Combrinck MI, De Jager CA. Prevalence of HIV-1 infection in an elderly rural population and association with impairment. AIDS. 2019;33(11):1765–1771. https://doi.org/10.1097/QAD.0000000000002257

37. Milanini B, Allen I, Paul R, Bahemana E, Kiweewa F, Nambuya A, et al. Frequency and predictors of HIV-related cognitive impairment in East Africa: The Africa Cohort Study (AFRICOS). J Acquir Immune Defic Syndr. 2020;83(2):157–164. https://doi.org/10.1097/QAI.0000000000002242

38. Mogambery JC, Dawood H, Wilson D, Moodley A. HIV-associated neurocognitive disorder in a KwaZulu-Natal HIV clinic: A prospective study. South Afr J HIV Med. 2017;18(1):732. https://doi.org/10.4102/sajhivmed.v18i1.732

39. Sacktor N, Nakassyja N, Okonkwo O, Skolasky RL, Robertson K, Musisi S, et al. Longitudinal neuropsychological test performance among HIV seropositive individuals in Uganda. J Neurovirol. 2013;19(1):48–56. https://doi.org/10.1007/s13365-012-0139-5

40. Patel SM, Thames AD, Arbid N, Panos SE, Castellon S, Hinkin CH. The aggregate effects of multiple comorbid risk factors on cognition among HIV-infected individuals. J Clin Exp Neuropsychol. 2013;35(4):421–434. https://doi.org/10.1080/13803395.2013.783000

41. Nightingale S, Joska JA, Winston A, Gisslén M, Barber T. Reader response: Global prevalence and burden of HIV-associated neurocognitive disorder: A meta-analysis. Neurology. 2020. Available from: https://n.neurology.org/content/reader-response-global-prevalence-and-burden-hiv-associated-neurocognitive-disorder-meta

42. Gisslen M, Price RW, Nilsson S. The definition of HIV-associated neurocognitive disorders: Are we overestimating the real prevalence? BMC Infect Dis. 2011;11, Art. #356. https://doi.org/10.1186/1471-2334-11-356

43. Habib AG, Yakasai AM, Owolabi LF, Ibrahim A, Habib ZG, Gudaji M. Neurocognitive impairment in HIV-1-infected adults in sub-Saharan Africa: A systematic review and meta-analysis. Int J Infect Dis. 2013;17(10):e820–e831. https://doi.org/10.1016/j.ijid.2013.06.011

44. Debalkie Animut AM, Sorrie MB, Birhanu YW, Teshale MY. High prevalence of neurocognitive disorders observed among HIV/AIDS in Southern Ethiopia: A cross-sectional study. PLoS ONE. 2019;14(3), e0204636. https://doi.org/10.1371/journal.pone.0204636

45. Mugendi AG, Kubo MN, Nyamu DG, Mwaniki LM, Wahome SK, Haberer JE. Prevalence and correlates of neurocognitive disorders among HIV patients on antiretroviral therapy at a Kenyan Hospital. Neurol Res Int. 2019;2019, Art. #5173289. https://doi.org/10.1155/2019/5173289

46. Granich R, Gupta S, Hersh B, Williams B, Montaner J, Young B, et al. Trends in AIDS deaths, new infections and ART coverage in the top 30 countries with the highest AIDS mortality burden; 1990-2013. PLoS ONE. 2015;10(7), e0131353. https://doi.org/10.1371/journal.pone.0131353

47. Molloy DW, Standish TI. A guide to the standardized Mini-Mental State Examination. Int Psychogeriatr. 1997;9(suppl 1):87–94; discussion 143–150. https://doi.org/10.1017/S1041610297004754

48. Power C, Selnes OA, Grim JA, McArthur JC. HIV dementia scale: A rapid screening test. J Acquir Immune Defic Syndr Hum Retrovirol. 1995;8(3):273–278. https://doi.org/10.1097/00042560-199503010-00008

49. Sumonu TA, Imarhiagbe F, Owolabi LF, Ogunrin OA, Komolafe MA, Ilesanmi OS. Cognitive functions in newly diagnosed patients with HIV infection in a tertiary health facility: Assessment using community screening interview for dementia. eNeurologicalSci. 2017;24(9):8–13. https://doi.org/10.1016/j.ensci.2017.10.001

50. Sacktor NC, Wong M, Nakasujja N, Skolasky RL, Selnes OA, Musisi S, et al. The International HIV Dementia Scale: A new rapid screening test for HIV dementia. AIDS. 2005;19(13):1367–1374.

51. De Almeida SM, Kumat R, Cherner M, Umlauf A, Ribeiro CE. Improving detection of HIV-associated cognitive impairment: Comparison of international HIV dementia scale and a brief screening battery. J Acquir Immune Defic Syndr. 2017;74(3):332–338. https://doi.org/10.1097/QAI.0000000000001224

52. Kakuba N, Menon JA, Franklin DR, Heaton RK, Hestad KA. Use of Western neuropsychological test battery in detecting HIV-associated neurocognitive disorders in Zambia. AIDS Behav. 2017;21(6):1717–1727. https://doi.org/10.1007/s10461-016-1443-5

53. Gandhi NS, Moxley RT, Creighton J, Roosa HV, Skolasky RL, Selnes OA, et al. Comparison of scales to evaluate the progression of HIV-associated neurocognitive disorder. HIV Ther. 2010;4(3):371–379. https://doi.org/10.2217/hiv.10.23

54. Tierney SM, Sheppard DP, Kordovski VM, Faytell MP, Avci G, Woods SP. A comparison of the sensitivity, stability, and reliability of three diagnostic schemes for HIV-associated neurocognitive disorders. J Neurovirol. 2017;23(3):404–421. https://doi.org/10.1007/s13365-016-0510-z

55. Akolo C, Royal W III, Cherner M, Okwuasaba K, Eyzaguirre L, Adebiyi R, et al. Neurocognitive impairment associated with predominantly early stage HIV infection in Abuja, Nigeria. J Neurovirol. 2014;20(4):380–387. https://doi.org/10.1007/s13365-014-0254-6

56. Gandhi NS, Moxley RT, Creighton J, Roosa HV, Skolasky RL, Selnes OA, et al. Comparison of scales to evaluate the progression of HIV-associated neurocognitive disorder. HIV Ther. 2010;4(3):371–379. https://doi.org/10.2217/hiv.10.23

57. Singh D, Joska JA, Goodkin K, Lopez E, Myer L, Paul RH. Normative scores for a brief neuropsychological battery for the detection of HIV-associated neurocognitive disorder (HAND) among South Africans. BMC Res Notes. 2010;3:28. https://doi.org/10.1186/1756-0500-3-28

58. Chetty V, Maharaj SS. Collaboration between health professionals in the era of antiretroviral therapy. J Assoc Nurses AIDS Care. 2013;24(2):166–175. https://doi.org/10.1016/j.jana.2012.04.005

59. Belete T, Medfu G, Yemiyamrew E. Prevalence of HIV associated neuro-deficit among HIV positive people in Ethiopia: A Cross-sectional study at Ayder Referral Hospital. Ethiop J Health Sci. 2017;27(1):67–76.

60. Hendershot CS, Stoner SA, Pantalone DW, Simoni JM. Alcohol use and antiretroviral adherence: Review and meta-analysis. J Acquir Immune Defic Syndr. 2009;52(2):180. https://doi.org/10.1097/QAI.0b013e3181b18b6e

61. Anderson SG, McCaul M, Khoo S, Wiesner L, Sacktor N, Joska JA, et al. The neurologic phenotype of South African patients with HIV-associated neurocognitive impairment. Neurol Clin Pract. 2020;10(1):15–22. https://doi.org/10.1212/CPJ.0000000000000687

62. Asiedu N, Kretchy I, Asampong E. Psycho-behavioral factors associated with neurocognitive performance among people living with HIV on antiretroviral therapy in Accra, Ghana. Afr Health Sci. 2020;20(2):487–596. https://doi.org/10.4314/ahs.v20i2.6

63. Hestad KA, Chinyama J, Anitha MJ, Ngoma MS, McCutchan JA, Franklin DR Jr, et al. Cognitive impairment in Zambians with HIV infection and pulmonary tuberculosis. J Acquir Immune Defic Syndr. 2019;80(1):110–117. https://doi.org/10.1097/QAI.0000000000001880

64. Joska JA, Westgarth-Taylor J, Myer L, Hoare J, Thomas KG, Combrinck M, et al. Characterization of HIV-associated neurocognitive disorders among individuals starting antiretroviral therapy in South Africa. AIDS Behav. 2011;15:1197–1203. https://doi.org/10.1007/s10461-010-9744-6

65. Joska JA, Fincham DS, Stein DJ, Paul RH, Seedat S. Clinical correlates of HIV-associated neurocognitive disorders in South Africa. AIDS Behav. 2010;14(2):371–378.

66. Mohamed AA, Oduor C, Kinyanjui D. HIV-associated neurocognitive disorders at Moi teaching and referral hospital, Eldoret, Kenya. BMC Neurol. 2020;20:280. https://doi.org/10.1186/s12883-020-01857-3

67. Nakasujja N, Skolasky RL, Musisi S, Allebeck P, Robertson K, Ronald A, et al. Depression symptoms and cognitive function among individuals with advanced HIV infection initiating HAART in Uganda. BMC Psychiatry. 2010;102, Art. #44. https://doi.org/10.1186/1471-244X-10-44

68. Nakku J, Kinyanda E, Hoskins S. Prevalence and factors associated with probable HIV dementia in an African population: A cross-sectional study of an HIV/AIDS clinic population. BMC Psychiatry. 2013;13, Art. #126. https://doi.org/10.1186/1471-244X-13-126

69. Njamnshi AK, Bissek AC, Ongolo-Zogo P, Tabah EN, Lekoubou AZ, Yepnjio FN, et al. Risk factors for HIV-associated neurocognitive disorders (HAND) in sub-Saharan Africa: The case of Yaoundé-Cameroon. J Neurol Sci. 2009;285(1–2):149–153.]https://doi.org/10.1016/j.jns.2009.06.043

70. Nweke MC, Akinpleu A, Ezema CI. Variation in spatio-temporal gait parameters among patients with HIV-related neurocognitive impairment. Indian J Physiother Occup Ther. 2019;13(4):186–191.

71. Nyamayaro P, Gouse H, Hakim J, Robbins RN, Chibanda D. Neurocognitive impairment in treatment-experienced adults living with HIV attending primary care clinics in Zimbabwe. BMC Infect Dis. 2020;20:383. https://doi.org/10.1186/s12879-020-05090-8

72. Tomita A, Ramlall S, Naidu T, Mthembu SS, Padayatchi N, Burns JK. Neurocognitive impairment risk among individuals with multiple drug-resistant tuberculosis and human immunodeficiency virus coinfection: Implications for systematic linkage to and retention of care in tuberculosis/human immunodeficiency virus treatment. J Nerv Ment Dis. 2019;207(4):307–310. https://doi.org/10.1097/NMD.0000000000000962

73. Tsegaw M, Andargie G, Alem G, Tareke M. Screening HIV-associated neurocognitive disorders (HAND) among HIV positive patients attending antiretroviral therapy in South Wollo, Ethiopia. J Psychiatr Res. 2017;85:37–41. https://doi.org/10.1016/j.jpsychires.2016.10.016

74. Yusuf AJ, Hassan A, Mamman AI, Muktar HM, Suleiman AM, Baiyewu O. Prevalence of HIV-associated neurocognitive disorder (HAND) among patients attending a tertiary health facility in northern Nigeria. J Int Assoc Provid AIDS Care. 2017;16(1):48–55. https://doi.org/10.1177/2325957414553839

AUTHORS:
Susan Taljaard[1,2]
Janine Adams[2,3]

AFFILIATIONS:
[1]Coastal Systems Research Group, Council for Scientific and Industrial Research, Stellenbosch, South Africa
[2]Institute for Coastal and Marine Research, Nelson Mandela University, Gqeberha, South Africa
[3]DSI/NRF Research Chair in Shallow Water Ecosystems, Department of Botany, Nelson Mandela University, Gqeberha, South Africa

CORRESPONDENCE TO:
Susan Taljaard

EMAIL:
staljaar@csir.co.za

DATES:
Received: 03 Sep. 2020
Revised: 23 Nov. 2020
Accepted: 29 Mar. 2021
Published: 29 Sep. 2021

HOW TO CITE:
Taljaard S, Adams J. Coastal management – working towards the UN's Decade of Ocean Science for Sustainable Development (2021–2030). S Afr J Sci. 2021;117(9/10), Art. #8857. https://doi.org/10.17159/sajs.2021/8857

ARTICLE INCLUDES:
☒ Peer review
☐ Supplementary material

DATA AVAILABILITY:
☐ Open data set
☐ All data included
☐ On request from author(s)
☐ Not available
☒ Not applicable

EDITORS:
Bettine van Vuuren
Sydney Moyo

KEYWORDS:
coastal management, scientific support, South Africa, UN Decade, ocean science for sustainable development

FUNDING:
Council for Scientific and Industrial Research (South Africa); National Research Foundation of South Africa (UID 84375)

Coastal management – working towards the UN's Decade of Ocean Science for Sustainable Development (2021–2030)

The UN declared 2021-2030 as the Decade of Ocean Science and identified research and technology priority areas to achieve the 2030 Sustainable Development Goals. We reviewed the current status of scientific support for coastal management in South Africa within the context of these priorities and found promising development. However, challenges for the next decade remain, such as rolling out pilot projects into sustainable, national-scale programmes, facilitating greater collaboration and coordination among scientific role players, and achieving long-term commitment and political will for dedicated financial support. Through our lens as natural scientists we focused on the ecological system and coupling with the social system; however scientific support on better characterisation and understanding of the dynamics within the social system is also critical as sustainable development relies heavily on the willingness of the social system to embrace and execute related policies.

Significance:

- The UN Decade of Ocean Science (2021–2030) sets research and technology priority areas to achieve the 2030 Sustainable Development Goals.
- We found promising development in scientific support for coastal management in South Africa.
- Future challenges include greater collaboration and coordination among scientific role players and long-term commitment and political will for dedicated financial support.

Introduction

The rich natural resources of coasts have attracted humans for centuries, through supplying resources that sustain subsistence livelihoods and offering easy access to maritime trade, safe recreational use, and a sense of place. Coastal areas are outstripping the hinterland in terms of development driven by rapid economic growth and coastward migration.[1] Human pressures are changing the character of these areas, most noticeably over the past five decades with the trend expected to continue in future[2], coinciding with global impacts such as climate change, sea level rise, and ocean acidification.

In response to growing threats, the concept of coastal zone management emerged in the early 1970s and was first captured in the USA's *Coastal Zone Management Act*[3], triggered by the visible human impacts on coasts. In the 1980s, the term 'integrated' was added when it became clear that effective management required inter-sectoral collaboration, and inclusion of socio-economic concerns. Numerous scientific studies have since contributed to the body of knowledge on integrated coastal management (ICM) and deepened the understanding of practice worldwide.[3,4]

Despite considerable investment, the implementation of ICM appears to have been deficient because degradation of coastal ecosystems continues.[2] Also, coastal systems are losing their resilience and becoming more vulnerable to environmental disasters. Further, the socio-economic factors responsible for this decline occur increasingly at scales greater than that of the typical (local) scope of ICM. To be effective, an adaptive management strategy[5] needs to be embraced, working across scales, to allow a broader perspective of the workings of coupled coastal social and ecological systems (SESs)[6].

Coastal management has received much attention in South Africa, especially post-1994 since adopting a power-sharing democratic political system. Environmental legislation adopted a pluralistic slant, embracing governance systems such as ecosystem-based management[7] and ICM[3] through laws such as the *National Environmental Management Act* (NEMA) and the *National Environmental Management Act: Integrated Coastal Management Act* (ICM Act). In 2014, South Africa published the first National Coastal Management Programme to assist with ICM implementation.[8,9] While activity-based legislation remains critical, confusion and conflicts often arise from poor understanding of interlinkages and inherent complexities.[10-12]

Despite a significant overhaul of its approach to coastal management, South Africa is struggling to implement the policies to achieve desired outcomes.[13-15] Many factors are responsible for this challenge, not least of which are the high degree of collaboration and resource capacity required to achieve these outcomes.[13] Other major challenges include the absence of political support, inadequate institutional capacity, financial limitations, uncertainties in roles and responsibilities, and limited civil society involvement in decision-making.[15] Potential solutions posed centre on broader representation in ICM initiatives, revitalisation of public interest, improvement of cooperative governance, increased funding to coastal management, and a greater commitment to a more deliberative and collaborative governance approach.[15] Globally, the ineffectual efforts to manage threats on oceans and coasts seriously hamper the ability of nations to achieve the goals of the 2030 UN Agenda for Sustainable Development. In response, the UN declared 2021–2030 as the Decade of Ocean Science for Sustainable Development (UN Decade), aimed at creating

a common framework to strengthen ocean and coastal management for human benefit.[16,17]

South Africa's National Biodiversity Assessment of 2018 considers the coast as including all '…terrestrial and marine ecosystem types with strong coastal affinities...'[18]. Landwards, this includes vegetation types described as purely coastal or having a coastal affinity. Seawards, it includes all ecosystem types influenced by land stretching as far seaward as the inner shelf, including embayments, river-influenced systems, and estuaries. Using this ecological demarcation, we reviewed the current status of scientific support for coastal management in South Africa in the context of the UN Decade's seven research and technology priorities through our lens as natural scientists, focussing on scientific support for management of the coastal, nearshore domain, with less attention on the offshore, deeper oceanic areas.

Role of scientific information and knowledge in coastal management

The coast can be viewed as a SES connected by the flow of ecosystem services (Figure 1).[19,20] Within this system, components can negatively impact each other, for example, environmental hazards can threaten the social component, while human interference can degrade the ecological component. Coastal management is depicted as the 'hands' protecting the system's integrity by coordinating interactions within the SES to sustain co-existence. Practising the scientific method generates reliable evidence-based information that can be used to understand processes within the SES. Scientific knowledge, therefore, is viewed as a knowledge type informing coastal policy development and management rather than advocating specific policy outcomes.[21,22]

The role of scientific information and knowledge in managing coastal SESs not only entails studying the characteristics and dynamics within each component but also determining how they interact with each other. Scientific studies should take into account place-based values and uncertainties relevant to policy imperatives. In South Africa, collaboration among coastal scientists and policymakers, especially at national and provincial tiers, has been growing since the promulgation of people-centred environmental legislation post-1994.[10,12] It is common practice for government departments to engage with scientists in research councils, institutions, and environmental consultancies, to assist with development of policies, guidelines and management strategies. With a few exceptions (e.g. fisheries), departments do not possess strong in-house scientific capacity and need external support. This, perhaps unintentionally, has led to a close collaboration between scientists and policymakers. It has not, however, lessened the serious challenges facing policy implementation and the growing need for science-based knowledge to lay the foundation for prudent coastal management in future.

UN's Priorities for Decade of Ocean Science for Sustainable Development (2021–2030)

The Decade is viewed as *'…a once in a lifetime opportunity to create a new foundation, across the science-policy interface, to strengthen the management of our oceans and coasts for the benefit of humanity…' with a vision of '…developing scientific knowledge, build infrastructure and foster partnerships for sustainable and healthy ocean and coasts…'* to deliver the underpinning scientific information to achieve the 2030 Sustainable Development Goals.[16,17] Role players, including scientists, are encouraged to go beyond 'business as usual' and develop innovative solutions for ocean and coastal sustainability. Such solutions should support the integration in the SES, and recognise trade-offs within.[23] The Decade aims to achieve six outcomes: oceans (and coasts) are (1) clean,(2) healthy and resilient, (3) predicted, (4) safe, (5) sustainably harvested and productive, and (6) transparent and accessible.[18] Although science is not the only route to success, seven research and technology priority areas are seen as necessary to achieving these outcomes on

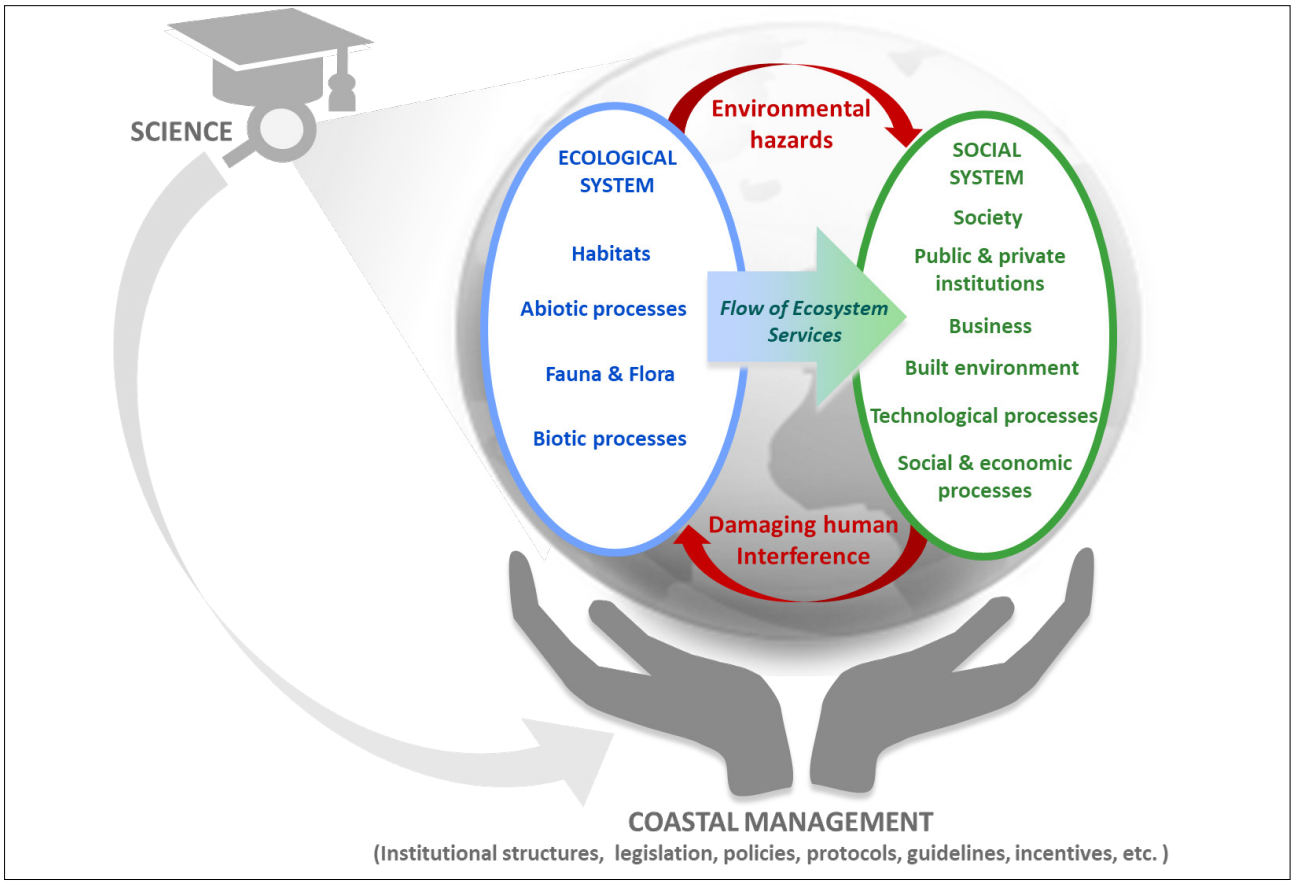


Figure 1: Conceptualisation of a coastal social-ecological system, as well as the role of coastal management and science.

both global and local scales, that is (1) geo-referenced digital atlases, (2) ocean observing systems (especially for major basins); (3) quantitative understanding of ocean ecosystems and their functioning as basis for management and adaptation; (4) ocean data and information portals; (5) integrated multi-hazard warning systems; (6) oceans in earth-system observation, research and prediction (prediction capabilities); and (7) capacity development and accelerated technology transfer, training and education.[16,17]

While the UN Decade aims to facilitate science–policy collaboration at the global ocean scale, its vision, goals and priorities can also guide nested, national-scale initiatives towards achieving the Sustainable Development Goals. We review the status of scientific support for coastal management in South Africa to see if we are on the right track.

Status of scientific support for coastal management in South Africa

Geo-referenced digital atlases

Internet-based digital atlases are increasingly accessible and of great value to role players interested in coastal issues[24], also for South Africa. The recent National Biodiversity Assessment has identified and mapped coastal and estuarine ecosystem types, as well as their ecological condition, ecosystem threat status and protection levels[18,25] (Table 1). Further, the *Green Book – Adapting Settlements for the Future* generated spatially referenced information on current and future settlement patterns and their vulnerability (Table 1). Land-use development planning is captured by municipalities in their Spatial Development Frameworks as mandated under the *Spatial Planning and Land Use Management Act (No. 16 of 2013)*, although these frameworks do not include the area seaward of the high-water mark. Spatial development planning in the coast and oceans is mostly sectoral (e.g. mining, fisheries, conservation), mandated to the responsible national department. With the promulgation of the *Marine Spatial Planning Act (No 16 of 2018)*, national government poses a mechanism for the coordination of spatial development planning within coasts and oceans, supported by a Marine Spatial Planning Tool being developed as part of the national Oceans and Coastal Information Management System (OCIMS) (Table 1) – an initiative between the Department of Forestry, Fisheries and the Environment and the Department of Science and Innovation (DSI) under Operation Phakisa, an initiative aimed at unlocking the economic potential of coastal and ocean resources. OCIMS's vision is to develop a national platform for the dissemination of data from authoritative sources. Also available is a coastal viewer, which provides spatial information pertaining to coastal public property and the coastal protection zone, as defined in the ICM Act.

Table 1: Summary of existing scientific initiatives aligned with the UN Decade's priorities supporting coastal management in South Africa

UN Decade priorities		EXISTING INITIATIVES IN SOUTH AFRICA
1	Geo-referenced digital atlases	South African National Biodiversity Assessment: Biodiversity GIS (http://bgis.sanbi.org/) The Greenbook (https://greenbook.co.za/) South Africa's National Oceans and Coastal Information Management System (OCIMS) (www.ocims.gov.za/)
2	Observing systems	OCIMS (www.ocims.gov.za/) Transnet National Ports Authority (TNPA) Wavenet (http://wavenet.csir.co.za/) Acoustic Tracking Array Platform (South African Institute for Aquatic Biodiversity) (www.saiab.ac.za/atap.htm) South African Weather Service (http://marine.weathersa.co.za/Observations_Home.html) Ecosystem long-term ecological research sampling stations (https://smcri.saeon.ac.za/SentinelSites/AlgoaBay) Beach water quality monitoring, City of Cape Town (http://www.capetown.gov.za/Explore%20and%20enjoy/nature-and-outdoors/our-precious-biodiversity/coastal-water-quality) Beach water quality monitoring, eThekwini Municipality (http://www.durban.gov.za/City_Services/water_sanitation/Water_Quality/Pages/ Beach_Water_Quality.aspx) National Pollution Laboratory, Walter Sisulu University (http://www.wsu.ac.za/waltersisulu/index.php/wsu-hosts-national-pollution-lab/)
3	Building understanding on ecosystem functioning	SANCOR (https://sancor.nrf.ac.za/) *African Journal of Marine Science* (www.ajol.info/index.php/ajms/) NRF Community of Practice (https://www.nrf.ac.za/division/funding/communities-practice)
4	Data and information portals	Southern African Data Centre for Oceanography, hosted by Marine Information Management System (http://data.ocean.gov.za/) National Ocean and Coastal Information Systems (OCIMS) (www.ocims.gov.za/) An interactive web-based information portal – CoastKZN (www.coastkzn.co.za/) South African Estuary Information System (https://saeis.saeon.ac.za/) Estuary botanical database hosted by Nelson Mandela University (http://opus.sanbi.org/handle/20.500.12143/6707) Barcode of Life Data System (BOLD) (http://boldsystems.org)
5	Prediction capabilities	OCIMS (https://www.ocims.gov.za/) SAWS (http://marine.weathersa.co.za/)
6	Multi-hazard warning systems	
7	Capacity building and technology transfer	NRF initiatives: Career Advancement, Thuthuka and Phuhlisa programmes (http://www.nrf.ac.za) NRF's Marine and Coastal Research funding instrument. Marine Research Plan (2014–2024) theme – with coastal and marine resources, society and development as one of the themes (https://www.nrf.ac.za/sites/default/files/documents/ Marine%20and%20Coastal%20Research%20Framework%202019.pdf) Marine and Coastal Educators Network, pilot testing inclusion of Marine Science as a subject at school (https://www.aquarium.co.za/blog/entry/new-marine-sciences-school-subject-for-grade-10-learners-in-2019#) One Ocean Hub funded by Global Challenge Research Fund (https://oneoceanhub.org/) Two Oceans Aquarium, Cape Town (https://www.aquarium.co.za/) Ushaka Marine World, Durban (https://ushakamarineworld.co.za/)

While progress has been made with the development of platforms to collate digital, spatially referenced environmental data and information relevant to coastal management, the challenge ahead is to expand the spatial coverage, and ensure its longevity through proper institutionalisation and allocation of dedicated resources. These platforms often experience a slow decline following the termination of initial development funding or the departure of enthusiastic champions who usually push their development. The production of countrywide digital coastal atlases cannot be achieved by government alone and will require formalised and funded partnerships with coastal scientists across tertiary and scientific institutions. Furthermore, cross-coordination and collaboration between platform custodians is critical, not only to ensure compatibility (e.g. standardisation of spatial mapping tools and protocols), but for quality control and cross-verification of spatial information to prevent overlap or conflicting messages.

Observing systems

The primary purpose of observing systems is to gather long-term data for the purposes of (1) building understanding on physical, chemical and biological processes and genetics; (2) characterising and understanding long-term changes and the role of global processes such as climate change; (3) understanding and quantifying specific and cumulative impacts of anthropogenic pressures; and (4) characterising and understanding the impacts of, and resilience to, extreme events.[26-28]

The South African coast is geomorphologically and ecologically complex and diverse, and supports at least 216 ecosystem types.[18] This complexity and diversity poses unique challenges for coastal ocean observing systems such as locating observing stations to develop representative pictures of regional and local conditions.[29] Operational management of health and safety associated with coastal and maritime activities such as maritime commerce, commercial fishing, recreation, search and rescue, and pollution management benefit greatly from coastal ocean observing systems.[29,30]

In South Africa, the national fisheries department has had the most success in sustaining long-term observing systems, linked to sustainable fisheries management initiatives.[31] Other observing systems were launched but were generally small in scope with unproven sustainability.[26] Transnet, the National Ports Authority, runs a network of buoys that collect real-time wave and wind data at major ports (Table 1), primarily to guide shipping operations but they have also proven valuable in assessing coastal vulnerability. Observing systems currently operational on the OCIMS platform include vessel movement tracking, harmful algal blooms, coastal sea state, and marine predators. The South African Institute for Aquatic Biodiversity manages the Acoustic Tracking Array Platform (Table 1), in partnership with the Canada-based global Ocean Tracking Network, using acoustic telemetry to track the movement of various species, especially those pertaining to resource management.[32]

Robotics (e.g. wave gliders) provides a cost-efficient survey technique that can function over large spatial and temporal scales.[33] The South African Weather Service (SAWS) have developed observing capacity in the country's coastal oceans using in-situ and remotely sensed tools (Table 1). Observations include sea-surface temperature and height, and surface currents. Remotely sensed satellite data (e.g. sea-surface temperatures, altimetry, ocean colour) are also increasingly applied in coastal observing systems through OCIMS (Table 1). Some metropolitan municipalities (e.g. City of Cape Town and eThekwini) support beach water quality observing systems, while a National Pollution Laboratory has been established at Walter Sisulu University to house coastal pollution observing systems (Table 1).

High-resolution observing systems are especially important in urban nodes, with them often situated along sheltered bays where anthropogenic pressures and management interventions are the most critical. Responding to this need, the Elwandle Coastal Node of the South African Environmental Observation Network (SAEON) launched the Algoa Bay Sentinel Site (Port Elizabeth and surrounds) in 2008, setting up several in-situ sensor moorings and acoustic receivers, estuary sensor moorings, and pelagic long-term research sampling stations (Table 1). Recently, the DSI, in conjunction with the National Research Foundation (NRF) and Nelson Mandela University (NMU), launched the shallow marine and coastal infrastructure initiative to provide instrumentation and platforms for the long-term observation of coastal regions and offshore islands.[34]

While there has been progress in expanding coastal observing systems, many of these are still in pilot phases, or are limited to specific areas, and will require extensive resources and commitment to roll out at the national scale. Remote sensing is a priority for future large-scale coastal and marine observing systems but cannot generate data on sub-surface processes. Deployment of fixed moorings and technologies such as autonomous floats and drones will be necessary to explore relationships among physical, chemical and biological processes and environmental change – critical for effective coastal management. Development of observing systems faces many challenges. Setting up instruments at sea, deploying vessels and ship's time are all costly. With dwindling financial resources, advances in observing systems will need to be prioritised and more effort made to coordinate use between role players such as government departments, public institutions such as the CSIR, SAWS, and SAEON, and private sector collaborators.

Building understanding on ecosystem functioning

In contrast to coastal ecosystem observing systems, South Africa has a rich history of coastal research undertaken by academic institutions, independent research organisations and science councils. National funding bodies such as the NRF and Water Research Commission have consistently supported research within the coastal space, as have international organisations including the Global Environmental Fund and United National Environmental Programme. Serving an advisory role, the South African Network for Coastal and Oceanic Research (SANCOR) is a non-statutory body established to generate and communicate knowledge and advice for policy development, use and management of coastal resources (Table 1) such as SEAChange, a successful research programme of which outcomes have been summarised in a special issue of the *African Journal of Marine Science* (Table 1). This journal also publishes papers on the conservation and management of living resources, relevant social science and governance, and new techniques. These knowledge sources provide a comprehensive bank of evidence-based information available to support government in formulating coastal policy and making ecologically prudent management decisions.

However, there is much still to be learnt, especially in the realm of global impacts on coastal ecosystems. With dwindling funding and growing competition for limited financial resources, interdisciplinary collaboration across institutions will be required to address such issues. In the past, large thematic research programmes created successful collaborative platforms such as the Benguela Ecology Programme, African Coelacanth Ecosystem Programme, and KwaZulu-Natal Bight Project.[25,35] Currently, the NRF's Marine and Coastal Research funding instrument together with the 2014–2024 Marine Research Plan, supports research and capacity development in coastal sectors (Table 1). The NRF's Community-of-Practice research programmes (Table 1) promote multidisciplinary SESs research, for example in Algoa Bay.[36]

Data and information portals

Since the 1960s, the Southern African Data Centre for Oceanography has curated, archived and disseminated marine and coastal data, now hosted through the Marine Information Management System (Table 1). At provincial level, CoastKZN is an interactive web-based information portal providing information on KwaZulu-Natal's coastal and estuarine environments (Table 1). SAEON has developed the South African Estuary Information System to ensure long-term data are archived and made accessible as a national asset (Table 1). NMU hosts an Estuary Botanical Database[37] that stores information on area coverage and species composition of habitat types in South African estuaries (Table 1).

While there has been progress in the development of web-based portals to support coastal management, their sustainability depends on the support of the scientific community and government's commitment

to long-term quality control, management and portal maintenance. Issues of data accessibility, ownership, quality, lack of awareness, and fragmentation in the acquisition and management of data across institutions often hamper the optimal use of resources. Frequently different formats and standards hamper efficiencies in locating, assembling and using available data, thus necessitating valuable time in acquiring even the most basic data and information, or recreating existing data sets. A solution is the development of portals with common web-based interfaces that provide fast and easy access to coastal data and information sources, both historical and real-time.[38] Global examples of open-source information on marine data include the Ocean Biogeographic Information System (OBIS) that stores data submitted through regional nodes (e.g. AfrOBIS).[39] The Barcode of Life Data System houses genetic information on South African species (Table 1), including coastal species.[28]

Prediction capabilities

Increasingly, conflicting uses of coastal resources are making uncoordinated silo-based (or sector-based) planning and management ineffective.[12] Integrative, scientific capabilities such as those of numerical and system dynamics modelling are required to provide support in addressing such inefficiencies, especially within high-use urban nodes[40], providing pro-active means of forecasting future scenarios for negotiation among all decision-makers[17,41]. Through OCIMS, South Africa is progressing with forecasting of key environmental variables such as winds, currents, and waves (i.e. the Coastal Operations at Sea Platform), as is SAWS, by applying technologies such as remote sensing, numerical modelling, and machine learning (Table 1).

Inclusion of emerging technologies, such as machine learning and artificial intelligence, holds great advantages for future predictive capabilities, while system dynamics modelling allows for contextualising responses to future scenarios in complex systems.[42,43] Concepts such as 'Building with Nature'[44] can be combined with coastal environmental sciences through disciplines such as engineering and architecture to align natural and engineered systems for future benefits. While these technologies and approaches are being applied to coastal planning at the project level and in forecasting change in specific variables, they have not yet been applied to resolving complex cross-disciplinary ecosystem dynamics, or in testing multi-use strategies for ICM. Advancements in future predictive capabilities in support of ICM will benefit greatly from such transdisciplinary collaboration.

Multi-hazard warning systems

Pro-active responses to hazards require early warning systems, preferably for responding simultaneously to multi-hazards associated with coastal erosion, flooding, sea level rise, storm surges and cyclones.[17,45] Effective early warning systems are not restricted to detection, monitoring and forecasting, but also include risk analysis, timeous dissemination of warnings, and preparation of appropriate emergency plans.[46]

Through OCIMS, progress is being made in providing a consolidated platform for detection, monitoring and forecasting of coastal hazards along South Africa's coast (Table 1). For example, the Coastal Flood Hazard Decision Support Tool aims to identify coastal areas at risk of flooding, and the Fisheries and Aquaculture Decision Support Tool will provide a capability for monitoring and assessing risk of harmful algal blooms. SAWS also have been developing early warning hazard support tools such as the Storm Surge Forecasting System (Table 1). However, key elements, such as timely dissemination of warnings and wide-scale formalisation of emergency response plans, need to be further developed. Also, early warning information on the African continent needs to address language barriers and consider available mobile phone technologies.[47,48]

Capacity development and technology transfer

The elements essential to achieve science support for coastal management include human potential, infrastructure, cooperation, resources, and adequate social conditions for successful R&D.[18] For the UN Decade objectives, countries like South Africa will need to strengthen their capacity and technological know-how through education and training.[49] At national level, the DSI and NRF Centres of Excellence and National Research Chairs aim to address such capacity building. Eight Research Chairs focus on coastal research, strengthening the ability of universities to produce high-quality research and capable students. Dedicated university institutes, such as the Institute for Coastal and Marine Research (at NMU), International Ocean Institute of Southern Africa (at University of the Western Cape) and Marine Research Institute (at University of Cape Town), also contribute to research capacity. NMU, in collaboration with CSIR, CapeNature, the Western Cape government, and the Department of Forestry, Fisheries and the Environment, also offers a longstanding training programme on estuary management. Numerous initiatives under the NRF also are at work to advance marine sciences (Table 1). The SEAmester, launched in 2016, is the first floating university introducing postgraduate students and interns to marine science.[50] South Africa has a long history of international collaboration on coastal and marine research – another important avenue for building capacity and ensuring high-quality research outputs. A recent initiative driving SES research is the One Ocean Hub funded by the Global Challenge Research Fund, which seeks to bridge current disconnections in law, science and policy and integrate governance frameworks to balance multiple ocean uses with conservation (Table 1).

Oceanariums, such as the Two Oceans Aquarium and uShaka Marine World (Table 1), play an important role in stimulating public awareness, as do non-government organisations such as the Wildlife & Environment Society of South Africa, WildOceans, Oceans Research, Knysna Basin Research, Sustainable Seas Trust, and the African Marine Waste Network. To promote coastal sciences within the education system, the Marine and Coastal Educators Network, a group affiliated with SANCOR, coordinates activities, and shares experiences and materials to energise upcoming generations in marine and coastal sciences. The inclusion of Marine Science as a subject at school level is an important advance that was pilot tested in 2019 (Table 1).

Reflection on the way forward

We reviewed the current status of scientific support for coastal management in South Africa within the context of the seven research and technology priority areas of the UN Decade. While good progress has been made in developing platforms for collating *spatially referenced environmental atlases*, the challenge ahead is to expand the spatial coverage of data and information, and ensure its longevity through proper institutionalisation and allocation of dedicated resources beyond the development funding phases, and to facilitate cross-coordination and collaboration between platform custodians. Similarly, development of coastal *observing systems* and *web-based data and information portals* for coastal management has also progressed but will require extensive effort and commitment to roll out at the national scale. With dwindling financial resources, advances in observing systems will need to be prioritised and more effort made to coordinate use between role players. *Building understanding on ecosystem functioning* can be strengthened through large-scale thematic research programmes necessary to provide a quantitative understanding of coastal ecosystems.

Future *predictive capabilities*, especially within high-use urban nodes, will require integrative capabilities to address cross-sectoral interdependencies and conflicts. Inclusion of emerging technologies, such as machine learning and artificial intelligence, hold great advantages in term of such predictive capabilities, while system dynamics modelling can assist with assessing the complexities inherent in coastal social-ecological systems and provide useful ways of visualising and testing system behaviour within different development scenarios. Solutions through concepts like 'Working with Nature' and 'Building with Nature' can also facilitate collaboration with engineering and architecture to achieve greater harmonisation between natural and engineered systems. Initiatives, like OCIMS, are giving the environmental aspects of *early warning systems* along South Africa's coast renewed attention. However other key elements of multi-hazard warning systems, such as timely dissemination of warnings and wide-scale formalisation of emergency response plans, need to be further developed, as early

warning information may not always reach those at highest risk. *Capacity development* is supported by many national efforts that should be upscaled to ensure transformation and quality research outputs. Both national and international networks are important in developing skills and expertise.

Promising development has been occurring in many of the UN Decade priority areas for research and technology. However, challenges for the next decade remain, such as rolling out developmental phases into sustainable, national-scale programmes, facilitating greater collaboration and coordination among role players, and achieving long-term commitment and political will for dedicated financial support for human and technological capacity development, both in the generation and dissemination of related scientific knowledge. Our lens focused on the ecological system and coupling with the social system, but scientific support on better characterisation and understanding of the dynamics within the social system is also critical. Ultimately, sustainable development within the coastal social-ecological system relies heavily on the willingness of the social system to embrace and execute related policies.

Acknowledgements

This article stems from a presentation entitled: 'Future science perspectives for coastal management' given at South Africa's first ICM Lekgotla (September 2019) under the *National Environmental Management: Integrated Coastal Management Act* on invitation of the then national Department of Forestry, Fisheries and the Environment. We thank the Council for Scientific and Industrial Research (CSIR) (South Africa) for financial support in producing this manuscript; the National Research Foundation of South Africa for funding J.A. through the support of the DSI/NRF Research Chair in Shallow Water Ecosystems (UID 84375); CSIR colleagues for valuable input; and Dr Lesley McGwynne for providing editorial input.

Competing interests

We have no competing interests to declare.

Authors' contributions

S.T.: Formulation of overarching construction and content of manuscript; experiential knowledge and analysis of available data and information; preparation and creation of manuscript. J.A.: Experiential knowledge and analysis of available data and information; significant contribution to preparation and creation of manuscript.

References

1. Neumann B, Vafeidis AT, Zimmermann J, Nicholls RJ. Future coastal population growth and exposure to sea-level rise and coastal flooding – A global assessment. PLoS ONE. 2015;10(3), e0118571. https://doi.org/10.1371/journal.pone.0118571
2. Mee L. Between the devil and the deep blue sea: The coastal zone in an era of globalisation. Estuar Coast Shelf Sci. 2012;96:1–8. https://doi.org/10.1016/j.ecss.2010.02.013
3. Stojanovic T, Ballinger RC, Lalwani CS. Successful integrated coastal management: Measuring it with research and contributing to wise practice. Ocean Coast Manag. 2004;47:273–298. https://doi.org/10.1016/j.ocecoaman.2004.08.001
4. Taljaard S, Slinger JH, Van der Merwe JH. Criteria for evaluating the design of implementation models for integrated coastal management. Coast Manag. 2011;39:628–655. https://doi.org/10.1080/08920753.2011.616670
5. Bornmann BT, Martin JR, Wagner FH, Wood G, Alegria J, Cunningham PG, et al., editors. Ecological stewardship: A common reference for ecosystem management. Amsterdam: Elsevier; 1999. p. 505–534.
6. Folke, C, Hahn T, Olsson P, Norberg J. Adaptive governance of social-ecological systems. Annu Rev Environ Resour. 2005;30:441–473. https://doi.org/10.1146/annurev.energy.30.050504.144511
7. Halpern BS, McLeod KL, Rosenberg AA, Crowder LB. Managing for cumulative impacts in ecosystem-based management through ocean zoning. Ocean Coast Manag. 2008;51:203–211. https://doi.org/10.1016/j.ocecoaman.2007.08.002
8. South African Department of Environmental Affairs (DEA). The National Coastal Management Programme of South Africa. Cape Town: DEA; 2014.
9. Taljaard S, Slinger JH, Van der Merwe JH. Dual adaptive cycles in implementing integrated coastal management. Ocean Coast Manag. 2013;84:23–30. https://doi.org/10.1016/j.ocecoaman.2013.07.003
10. Taljaard S, Van Niekerk L, Weerts SP. The legal landscape governing South Africa's coastal marine environment – Helping with the 'horrendogram'. Ocean Coast Manag. 2019;178:104801. https://doi.org/10.1016/j.ocecoaman.2019.05.003
11. Cochrane KL, Joyner J, Sauer WHH, Swan J. An evaluation of the Marine Living Resources Act and supporting legal instruments as a framework for implementation of an ecosystem approach to fisheries in South Africa. Afr J Mar Sci. 2015;37:437–456. https://doi.org/10.2989/1814232X.2015.1100682
12. Taljaard S, Slinger JH, Morant PD, Theron AK, Van Niekerk L, Van der Merwe JH. Implementing integrated coastal management in a sector based governance system. Ocean Coast Manag. 2012;67:39–53. https://doi.org/10.1016/j.ocecoaman.2012.06.003
13. Goble BJ, Lewis M, Hill TR, Phillips MR. Coastal management in South Africa: Historical perspectives and setting the stage of a new era. Ocean Coast Manag. 2014;91:32–40. https://doi.org/10.1016/j.ocecoaman.2014.01.013
14. Von den Heyden S, Lukey P, Celliers L, Prochazka K, Lombard AT. Science to policy – Reflections on the South African reality. S Afr J Sci. 2016;112(11/12), Art. #a0183. https://doi.org/10.17159/sajs.2016/a0183
15. Sowman M, Malan N. Review of progress with integrated coastal management in South Africa since the advent of democracy. Afr J Mar Sci. 2018;40(2):121–136. https://doi.org/10.2989/1814232X.2018.1468278
16. Intergovernmental Oceanographic Committee of UNESCO. The science we need for the ocean we want: The United Nations Decade of Ocean Science for Sustainable Development (2021-2030). (English) IOC Brochure 2018-7 (IOC/BRO/2018/7 Rev) [webpage on the Internet]. c2019 [cited 2020 Nov 20]. Available from: https://www.oceandecade.org/resource/56/Brochure-The-Science-We-Need-for-the-Ocean-We-Want
17. Ryabinin V, Barbière J, Haugan P, Kullenberg G, Smith N, McLean C, et al. The UN Decade of Ocean Science for Sustainable Development. Front Mar Sci. 2019;6:470. https://doi.org/10.3389/fmars.2019.00470
18. Harris LR, Sink KJ, Skowno AL, Van Niekerk L, editors. South African National Biodiversity Assessment 2018: Technical report. Volume 5: Coast. Pretoria: South African National Biodiversity Institute; 2019. http://hdl.handle.net/20.500.12143/6374
19. Resilience Alliance. Assessing and managing resilience in social-ecological systems: Supplementary notes to the practitioners workbook. Vol. 2 (version 1.0). Resilience Alliance; 2007.
20. Bennett EM, Peterson GD, Gordon LJ. Understanding relationships among multiple ecosystem services. Ecol Lett. 2009;12:1394–1404. https://doi.org/10.1111/j.1461-0248.2009.01387.x
21. Gluckman P. Policy: The art of science advice to government. Nature. 2014;507:163–165. https://doi.org/10.1038/507163a
22. Gluckman P. The science-policy interface. Science. 2016;353:969. https://doi.org/10.1126/science.aai8837
23. Claudet J, Bopp L, Cheung WWL, Devillers R, Escobar-Briones E, Haugan P, et al. A roadmap for using the UN Decade of Ocean Science for Sustainable Development in support of science, policy, and action. One Earth. 2019;2(1):34–42. https://doi.org/10.1016/j.oneear.2019.10.012
24. Maelfait H, Belpaeme K. The Belgian Coastal Atlas: Moving from the classic static atlas to an interactive data-driven atlas. J Coast Conserv. 2010;14:13–19. https://doi.org/10.1007/s11852-009-0076-5
25. Van Niekerk L, Adams JB, Lamberth SJ, MacKay CF, Taljaard S, Turpie JK, et al., editors. South African National Biodiversity Assessment 2018: Technical report. Volume 3: Estuarine realm. Pretoria: South African National Biodiversity Institute; 2019. http://hdl.handle.net/20.500.12143/6373

26. Moloney CL, Shillington FA. Progress towards marine ecosystem observing systems in South Africa. S Afr J Sci. 2007;103:301–305.

27. Cocquempot L, Delacourt C, Paillet J, Riou P, Aucan J, Castelle B, et al. Coastal ocean and nearshore observation: A French case study. Front Mar Sci. 2019;6:324. https://doi.org/10.3389/fmars.2019.00324

28. Da Silva JM, Willows-Munro S. Review of over a decade of DNA barcoding in South Africa: A faunal perspective. Afr Zool. 2016;51(1):1–12. https://doi.org/10.1080/15627020.2016.1151377

29. Kite-Powell HL, Colgan CS. The potential economic benefits of coastal ocean observing systems: The Gulf of Maine. Silver Spring, MD: US Department of Commerce, National Oceanic and Atmospheric Administration, Office of Program Planning and Integration; 2001.

30. De Vos M, Rautenbach C. Investigating the connection between metocean conditions and coastal user safety: An analysis of search and rescue data. Safety Sci. 2019;117:217–228. https://doi.org/10.1016/j.ssci.2019.03.029

31. Hutchings L, Roberts MR, Verheye HM. Marine environmental monitoring programmes in South Africa: A review. S Afr J Sci. 2009;105:94–102. https://doi.org/10.4102/sajs.v105i3/4.54

32. Cowley PD, Bennett RH, Childs A-R, Murray TS. Reflection on the first five years of South Africa's Acoustic Tracking Array Platform (ATAP): Status, challenges and opportunities. Afr J Mar Sci. 2017;39:363–372. https://doi.org/10.2989/1814232X.2017.1399927

33. Swart S, Zietsman JJ, Coetzee JC, Goslett DG, Hoek A, Needham D, et al. Ocean robotics in support of fisheries research and management. Afr J Mar Sci. 2016;38:525–538. https://doi.org/10.2989/1814232X.2016.1251971

34. South African Department of Science and Technology (DST), Nelson Mandela University (NMU). SAEON & SAIAB to lead new research initiative to reduce SA's coastline vulnerability to global change [document on the Internet]. c2018 [cited 2020 Nov 20] Available from: https://www.nrf.ac.za/sites/default/files/documents/09%20NRF%20SMM%20V1%20Issue2%20SAEON%20%26%20SAIAB.pdf

35. Fennessy ST, Roberts MJ, Paterson AW. A brief overview of the ACEP Project: Ecosystem processes in the KwaZulu Natal Bight. Afr J Mar Sci. 2016;38(suppl):S1–S6. https://doi.org/10.2989/1814232X.2016.1141116

36. Dorrington RA, Lombard AT, Bornman TG, Adams JB, Cawthra HC, Deyzel SH, et al. Working together for our oceans: A marine spatial plan for Algoa Bay, South Africa. S Afr J Sci. 2018;114(3–4):1–6. https://doi.org/10.17159/sajs.2018/a0247

37. Adams JB, Veldkornet D, Tabot P. Distribution of macrophyte species and habitats in South African estuaries. S Afr J Bot. 2016;107:5–11. https://doi.org/10.1016/j.sajb.2016.08.001

38. Palazov A, Stefanov A, Marinova V, Slabakova V. Marine web portal as an interface between users and marine data and information sources. EGUGA. 2012:7083.

39. Yemane D, Kirkman SP, Samaai T. Use of openly available occurrence data to generate biodiversity maps within the South African EEZ. Afr J Mar Sci. 2020;42:109–121. https://doi.org/10.2989/1814232X.2020.1737573

40. Iglesias I, Avilez-Valente P, Luís Pinho J, Bio A, Manuel Vieira J, Bastos L, et al. Numerical modeling tools applied to estuarine and coastal hydrodynamics: A user perspective. In: Coastal and marine environments – Physical processes and numerical modelling. London: Intechopen; 2019. p. 20. https://doi.org/10.5772/intechopen.85521

41. Wilkin J, Rosenfeld L, Allen A, Baltes R, Baptista A, He R, et al. Advancing coastal ocean modelling, analysis, and prediction for the US Integrated Ocean Observing System. J Oper Oceanogr. 2017;10(2):115–126. https://doi.org/10.1080/1755876X.2017.1322026

42. Stock A, Haupt AJ, Mach ME, Micheli F. Mapping ecological indicators of human impact with statistical and machine learning methods: Tests on the California coast. Ecol Inform. 2018;48:37–47. https://doi.org/10.1016/j.ecoinf.2018.07.007

43. Chenar SS, Deng Z. Development of artificial intelligence approach to forecasting oyster norovirus outbreaks along Gulf of Mexico coast. Environ Int. 2018;111:212–223. https://doi.org/10.1016/j.envint.2017.11.032

44. Vikolainen, V, Bressers, H, Lulofs, K. A shift toward building with nature in the dredging and port development industries: Managerial implications for projects in or near Natura 2000 areas. Environ Manag. 2014;54:3–13. https://doi.org/10.1007/s00267-014-0285-z

45. Haigh R, Amaratunga D, Hemachandra K. A capacity analysis framework for multi-hazard early warning in coastal communities. Procedia Eng. 2018;212:1139–1146. https://doi.org/10.1016/j.proeng.2018.01.147

46. Rogers D, Tsirkunov V. Implementing hazard early warning systems. GFDRR WCIDS Report 11-03. Report prepared for Global Facility for Disaster Reduction and Recovery, Washington, USA [document on the Internet]. c2011 [cited 2020 Aug 26]. Available from: www.preventionweb.net/files/24259_implementingearlywarningsystems1108.pdf

47. South African Department of Environmental Affairs (DEA). Climate information and early warning system for supporting the disaster risk reduction and management sector in South Africa under future climates. Munzhedzi SM, Khavhagali VP, Midgley GM, De Abreu P, Scorgie S, Braun M, et al., editors. Long-term Adaptation Scenarios Flagship Research Programme. Pretoria: DEA; 2016.

48. Chaves JM, De Cola T. Public warning applications: Requirements and examples. In: Wireless Public Safety Networks 3. Amsterdam: Elsevier; 2017. p. 1–18. https://doi.org/10.1016/B978-1-78548-053-9.50001-9

49. Santoro F, Santin S, Scowcroft G, Fauville G, Tuddenham P. Ocean literacy for all: A toolkit (IOC Manuals and Guides, 80). Paris: UNESCO; 2017.

50. Ansorge IJ, Brundrit G, Brundrit J, Dorrington R, Fawcett S, Gammon D, et al. SEAmester – South Africa's first class afloat. S Afr J Sci. 2016;12(9/10):21–24. https://doi.org/10.17159/sajs.2016/a0171

AUTHORS:
Calvin Wells[1]
Justin Pringle[1]
Derek Stretch[1]

AFFILIATION:
[1]Environmental Fluid Mechanics Lab, University of KwaZulu-Natal, Durban, South Africa

CORRESPONDENCE TO:
Justin Pringle

EMAIL:
pringlej@ukzn.ac.za

DATES:
Received: 11 Dec. 2020
Revised: 07 Apr. 2021
Accepted: 22 Apr. 2021
Published: 29 Sep. 2021

HOW TO CITE:
Wells C, Pringle J, Stretch D. Cold water temperature anomalies on the Sodwana reefs and their driving mechanisms. S Afr J Sci. 2021;117(9/10), Art. #9304. https://doi.org/10.17159/sajs.2021/9304

ARTICLE INCLUDES:
☒ Peer review
☐ Supplementary material

DATA AVAILABILITY:
☐ Open data set
☐ All data included
☒ On request from author(s)
☐ Not available
☐ Not applicable

EDITORS:
Yali Woyessa
Jennifer Fitchett

KEYWORDS:
coral reefs, Sodwana reefs, hydrodynamics, temperature anomalies, coral bleaching

FUNDING:
eThekwini Chair in Civil Engineering (UKZN and eThekwini Municipality), South African National Research Foundation (UID: 127361)

Cold water temperature anomalies on the Sodwana reefs and their driving mechanisms

The Sodwana reef system experiences short-term temperature fluctuations that may provide relief from bleaching and be crucial in the future survival of the system. These temperature fluctuations are best described as cold water temperature anomaly events that occur over a period of days and cause a drop in temperature of a few degrees on the reef. We explored the statistical link between the temperature anomalies and the regional hydrodynamics to elucidate the driving mechanisms of the temperature anomalies around Sodwana. Temperature measurements taken between 1994 and 2015 on Nine-Mile Reef at Sodwana show that temperature anomalies occur on average three times per year at Sodwana and predominantly during the summer months. A conditional average of altimetry data at the peak of the temperature anomalies showed the emergence of a negative sea surface height (SSH) anomaly pattern and associated cyclonic eddy just offshore of the Sodwana region. The cyclonic eddies associated with the temperature anomalies originate on the southwestern edge of Madagascar and migrate westwards until they interact with the African coastline at Sodwana. Instantaneous altimetry SSH fields over the 21-year period were cross-correlated to the conditionally averaged SSH field within a 2° region around Sodwana. It was found that 33% of the temperature anomalies at Sodwana were not associated with the presence of cyclonic eddy systems. This finding suggests that an offshore cyclonic eddy interacting with the shelf is not the sole driving mechanism of the temperature anomalies.

Significance:

- Cold water temperature anomalies that occur at Sodwana are believed to play an important role in the sustainability of the coral reefs; however, their driving mechanisms are not yet well understood. We identified the annual frequency and seasonal occurrence of the temperature anomalies at Sodwana. Conditionally averaged sea surface heights delineated regional hydrodynamic patterns associated with the temperature anomalies. These hydrodynamic patterns could be a possible driving mechanism to be explored in future research. Previous research associated these temperature anomalies with a single regional hydrodynamic pattern around Sodwana; however, additional hydrodynamic patterns associated with the temperature anomalies have been found based on the length of the data sets used in this study.

Introduction

Coral reefs play a strong role in sustaining complex nearshore ecosystems. However, these coral reef systems are acutely sensitive to their surrounding ambient conditions. For example, temperature is vital to the sustainability and health of the coral reef system.[1] Coral reefs are commonly located in tropical environments and thrive in warm water that ranges between 18 °C and 30 °C.[2] However, corals become stressed when exposed to elevated temperatures for extended periods.[3,4] During these periods, the corals expel their symbiotic dinoflagellate algae, which is known as coral bleaching.[4,5] Severe bleaching may result in permanent reef damage and coral mortality from which the reef may not recover.[4,6]

Coral bleaching events at a global scale have been increasing in recent decades.[7,8] It is likely that these events will increase in frequency due to global climate change.[9,10] However, coral reefs located in certain high-latitude regions have a high bleaching refuge potential and are more suited to survive the projected rise in global ocean temperatures.[11] This is because these coral reefs are located in regions where there is strong ocean mixing.[11] An example is the Sodwana coral reef system located on the northern KwaZulu-Natal coast, South Africa. Sodwana is situated near the origin of the Agulhas Current in an area associated with complex hydrodynamic interactions with the continental shelf.[12] Sodwana experiences short-term temperature fluctuations that may provide relief from bleaching and be crucial in the future survival of the system.[13,14] These temperature fluctuations are best described as cold water temperature anomaly events that occur over a period of days and cause a drop in temperature of a few degrees on the reef. Roberts et al.[15] and Morris et al.[16] investigated one specific temperature anomaly event and suggested that the driving mechanism was associated with mesoscale cyclonic eddies interacting with the African coastline and shelf. For example, the presence of a cyclonic eddy near the shelf has been observed to be associated with an offshore deflection of the Agulhas Current that pulls inshore water away from the coast resulting in upwelling on the shelf.[15,16] It is still unclear if the time sequence of the temperature anomalies is always the same and whether all temperature anomalies observed at Sodwana are associated with cyclonic eddy upwelling. It is possible that there is a more complex combination of hydrodynamic factors driving the temperature anomalies. Understanding these temperature anomalies and their driving mechanisms is important as it provides a basis for predicting the coral reef health and survival along the Sodwana coastline. Furthermore, if these cold water temperature anomalies are driven by upwelling, as previous research suggests, it means that these events could also provide an influx of nutrients onto the reef.[17]

In this study, we aimed to elucidate the driving mechanisms of the cold water temperature anomalies around Sodwana, considering all temperature anomalies identified between 1994 and 2015.

Method

Case study site

The Sodwana reef system is one of the most southerly coral reefs in the world and comprises several marginal coral reef patches in the Delagoa Bioregion of South Africa.[18] The reefs are non-accretive and are rich in coral biodiversity and cover.[11,19] These reefs are world renowned and have become an important diving destination internationally. Figure 1 shows the location and bathymetry of the Sodwana region and reef system. The continental shelf in this region is narrow and is characterised by steep gradients and submarine canyons on the continental slope.[18] Sodwana is situated near the origin of the warm Agulhas Current, which raises the ocean temperatures in the region into the range that is conducive for coral growth.[18] The region also experiences high mesoscale eddy kinetic energy, and these eddies can interact with the complex shelf topography.[12]

Many coral reef systems located at higher latitudes have already begun to experience the effects of warming events associated with climate change.[11] Sodwana has shown resilience to these warming and bleaching events, which has been attributed to strong ocean mixing and short-term cold water anomalies.[13] This makes Sodwana an ideal case study site to investigate the complex hydrodynamic features that drive these temperature anomalies and how they influence the sustainability of higher latitude coral reefs.

Data

Measured temperature data

Hourly temperature data measured at Sodwana between 1994 and 2015 were used to investigate the short-term temperature trends on the reefs. The temperature measurements were obtained using an individual (self-contained) Star-Oddi mini underwater temperature recorder with an accuracy of ±0.05 °C, deployed at a depth of 18 m on Nine-Mile Reef (as per Figure 1).[15]

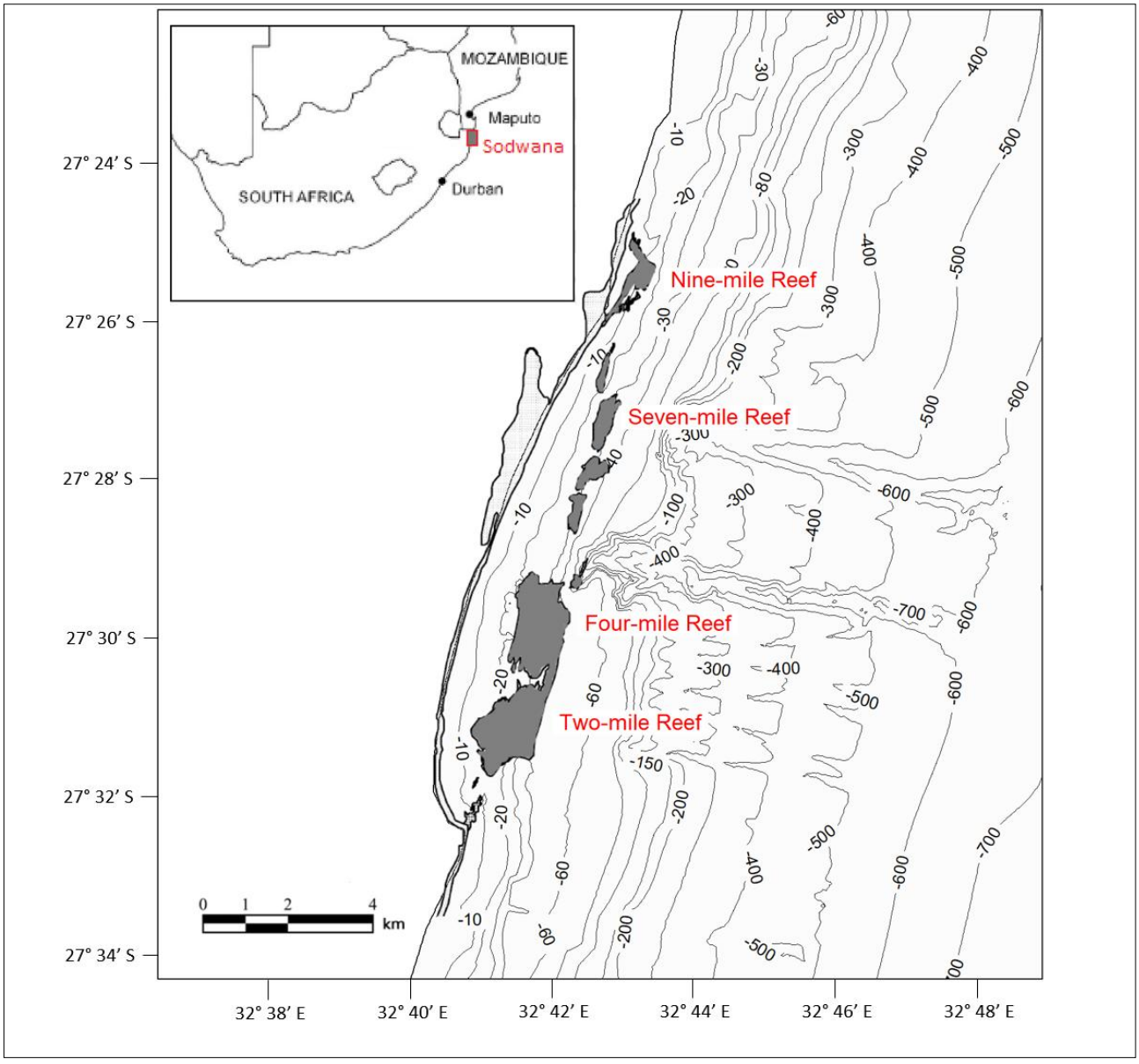


Figure 1: Sodwana Bay location and bathymetry.

Satellite data

Satellite altimetry data were used to link regional hydrodynamic patterns to the temperature anomalies observed at Sodwana. Satellite altimetry data are available from the Copernicus Marine Environment Monitoring Service which provides gridded sea surface height (SSH) data and derived geostrophic currents. The data are available globally with a horizontal resolution of 0.25° on a daily time step between 1993 and 2019.

Global eddy trajectory data

Global mesoscale eddy trajectory data were used to investigate the origins and temporal effects of the cyclonic eddies associated with the temperature anomalies at Sodwana. Global mesoscale eddy trajectory data were derived using an autonomous eddy identification and tracking procedure on the SSH satellite altimetry data.[20,21] The procedure is a modified version of the method presented by Williams et al.[22] The global eddy trajectory data are available from: http://cioss.coas.oregonstate.edu/eddies/.

Temperature analysis

The temperature on the Sodwana reefs fluctuates over a range of time scales.[12] The scales range from short hourly scales influenced by the tides to longer scales influenced by regional hydrodynamics and seasonal climatology. The temperature data were filtered to delineate the different time scales based on their frequency.[23] Internal waves have been observed to break on continental shelf slopes at a tidal frequency and have been identified as one possible mechanism for pushing colder water onto the shelf.[24,25] This becomes an important mechanism for regulating the reef temperature on a diurnal time scale. The temperature data were band pass filtered for periods between 10 h and 14 h to investigate temperature variability at the dominant local tidal frequency. The cold water temperature anomalies observed at Sodwana occur over a longer duration and regulate the reef temperature on a time scale of days.[15,16] Therefore, seasonal temperature variations were removed using a high pass filter for periods longer than a month. The remaining temperature signal was then used to identify the cold water temperature anomalies with time scales longer than the tidal periods but much shorter than seasonal variations. The temperature anomalies were identified using a local minimum threshold of -2 °C, which delineated temperature anomalies significantly different from the seasonal mean.

Conditional averaging

The satellite altimetry data were statistically analysed to link regional hydrodynamic features to the observed temperature anomalies. The SSH and geostrophic current fields at the peak of the temperature anomalies were conditionally averaged. Conditional averaging is a technique used when investigating statistically significant patterns in turbulent flow structures.[26,27] The conditional average was used to identify regional hydrodynamic patterns during the temperature anomalies that are significantly different from the unconditional mean. The conditional averaging of the SSH is defined as:

$$\overline{H(x, t_{i+\tau} \mid \Theta)} = \frac{\sum_{i=1}^{n_\theta} H(x, t_{i+\tau} \mid \Theta)}{n_\theta} \qquad \text{Equation 1}$$

where $\bar{H}$ is the conditional ensemble averaged SSH field, H is the instantaneous SSH field, x is the position vector, t_i represents the time at the peak of an individual temperature anomaly, τ is the time lag around the peak, Θ is the level of conditional threshold used for identifying temperature anomalies and n_θ presents the number of temperature anomalies identified with the threshold Θ. The U and V geostrophic velocity components were also conditionally averaged and used to calculate the conditionally averaged geostrophic current velocity fields. Time lags were considered from 15 days before until 10 days after the peak of the temperature anomalies.

The Kolmogorov–Smirnov statistic[28] was calculated for peak anomaly values within each conditional SSH data set for the various temperature thresholds. This was done to access if the conditional averaged SSH data sets for various thresholds were significantly different from the unconditional SSH data set at a significance level of 5%.

Cross-correlation

The instantaneous SSH fields over the 21 years were cross-correlated with the conditionally averaged SSH field to assess the fit of the individual anomalies to the conditional average.[29] The cross-correlation at each SSH timestep can be calculated with

$$p_{xy} = \frac{\overline{\overline{H(x, t_i)} \cdot H(x, t_i)} - \mu_{\bar{H}}\mu_H}{\sigma_{\bar{H}}\sigma_H} \qquad \text{Equation 2}$$

where $\overline{\bar{H} \cdot H}$ is the ensemble average of the dot product between the conditionally averaged ensemble SSH field and the instantaneous SSH field at a point in time t_i, x is the position vector, $\mu_{\bar{H}}$ and μ_H are the means of the conditionally averaged SSH field and instantaneous SSH field, respectively, and $\sigma_{\bar{H}}$ and σ_H are their standard deviations. The cross-correlation coefficients of the individual temperature anomalies were also used to identify any anomalies that do not exhibit the hydrodynamic patterns associated with the conditional average.

Cyclonic eddy tracking

The global cyclonic eddy trajectories were filtered to identify and track the mesoscale cyclonic eddies.[20] The identified cyclonic eddy trajectories were used to investigate the origin and temporal evolution of the eddies, and to link them to the observed mesoscale eddies linked to temperature anomalies.

Results

Temperature analysis

Cold water temperature anomalies occur throughout the temperature records on Nine-Mile Reef. Roberts et al.[15] identified notable temperature anomalies during the year 2004. Figure 2 presents a time series of the measured temperatures at Nine-Mile Reef during the year 2004 as an example to highlight the temperature anomalies in the data.

Five temperature anomalies can be clearly identified over the duration of the year: two occurring in February, one in April and two in December. Seasonal temperature variations can also be observed with higher temperatures in summer and lower temperatures in winter. Minimum temperatures observed during the summer anomalies were generally lower than the minimum temperatures recorded during winter.

Figure 3 shows an example of the temperature signal filtered at the tidal frequency for the month of February in 2004. This month was selected as an example period as it covers a spring-neap tidal cycle and two temperature anomalies which peaked on 14 February (T1) and 23 February (T2). The predicted water levels near Sodwana have been overlaid onto the time series plot.

The temperature signal filtered at a tidal frequency decreases with the incoming tide and increases with the outgoing tide. The tidal temperature fluctuations vary in amplitude between 0.1 °C and 1 °C. The increase in the temperature signal amplitude seems to amplify over the duration of the temperature anomaly. The variability in amplitude does not appear to coincide with the spring-neap tide cycle.

Figure 4 presents a time series of the measured temperature signal at Nine-Mile Reef with the tidal and seasonal temperature variance removed. Figure 4 also shows the temporal positions of the temperature anomalies exceeding the -2 °C threshold. These data were used to identify 63 temperature anomalies over the 21 years' of measured temperature data.

The temperature anomalies last on average for 6 days from start to finish with the longest event (peak of anomaly occurred on 25 December 2012) lasting 10 days and the shortest (peak of anomaly occurred on 6 January 2002) lasting only 3 days. The data indicate a lack of temperature anomalies between 1997 and 1999 and between 2007 and 2009.

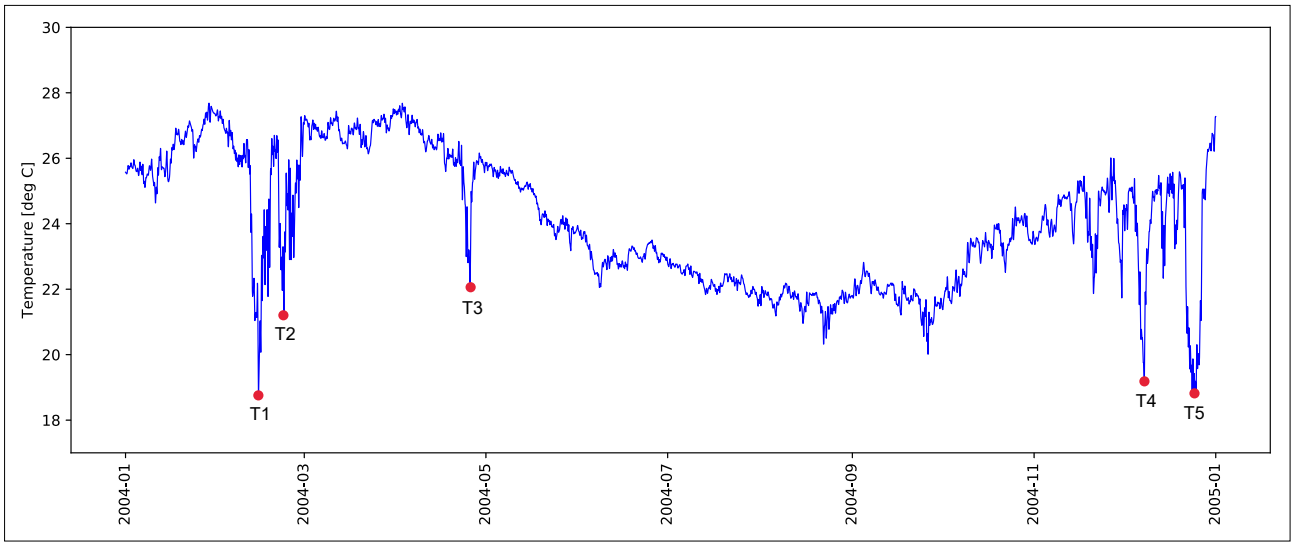


Figure 2: Time series of measured hourly temperature at Nine-Mile Reef during 2004. The individual temperature anomaly peaks are indicated by red dots.

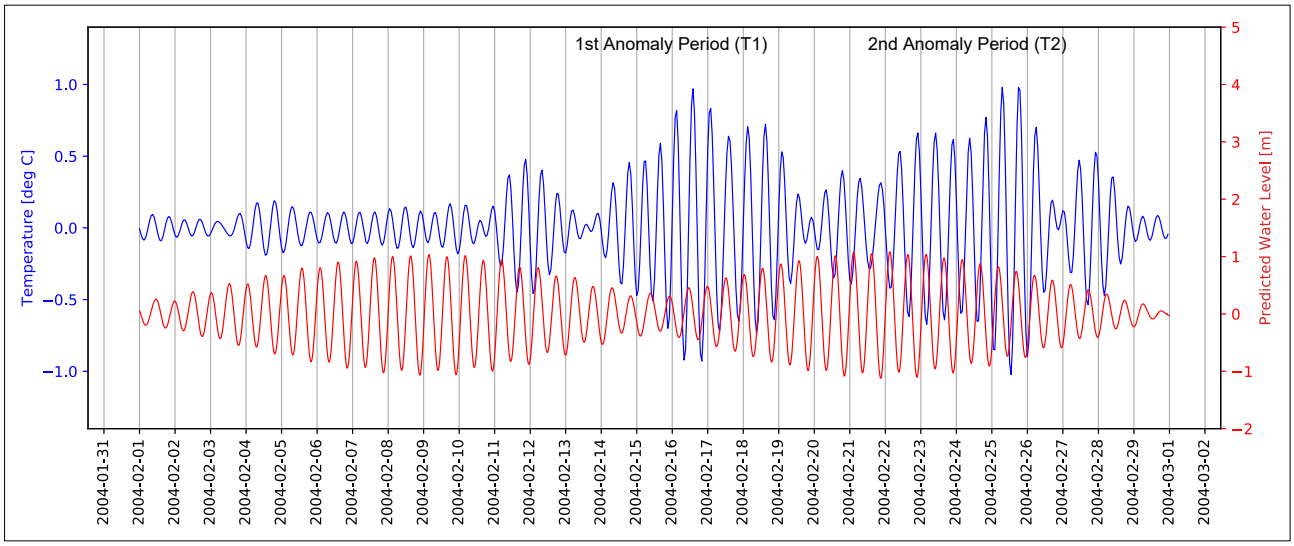


Figure 3: Time series of tidal temperature signal. The temperature signal is presented in blue and the predicted water levels are presented in red. The first anomaly period (T1) refers to the temperature anomaly that peaked on 14 February 2004 and the second anomaly period (T2) refers to the temperature anomaly that peaked on 23 Ferbruary 2004 (see Figure 2).

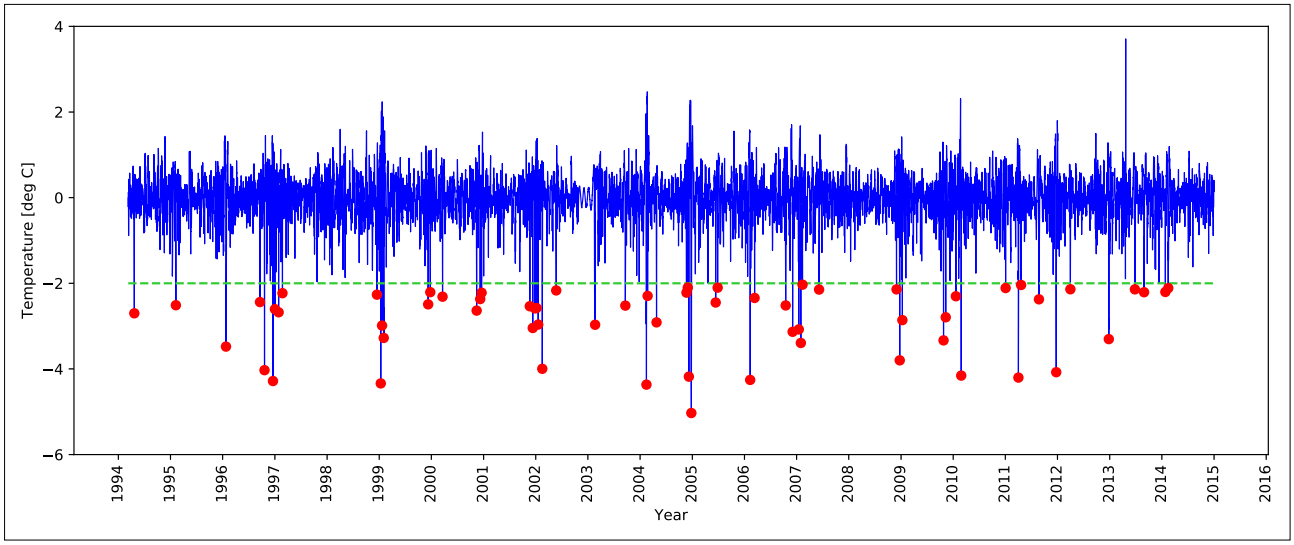


Figure 4: Filtered temperature signal on Nine-Mile Reef with the tidal and seasonal variance removed. Temporal positions of the temperature anomalies are denoted by red dots. The green dashed line presents the -2 °C threshold used for temperature anomaly identification.

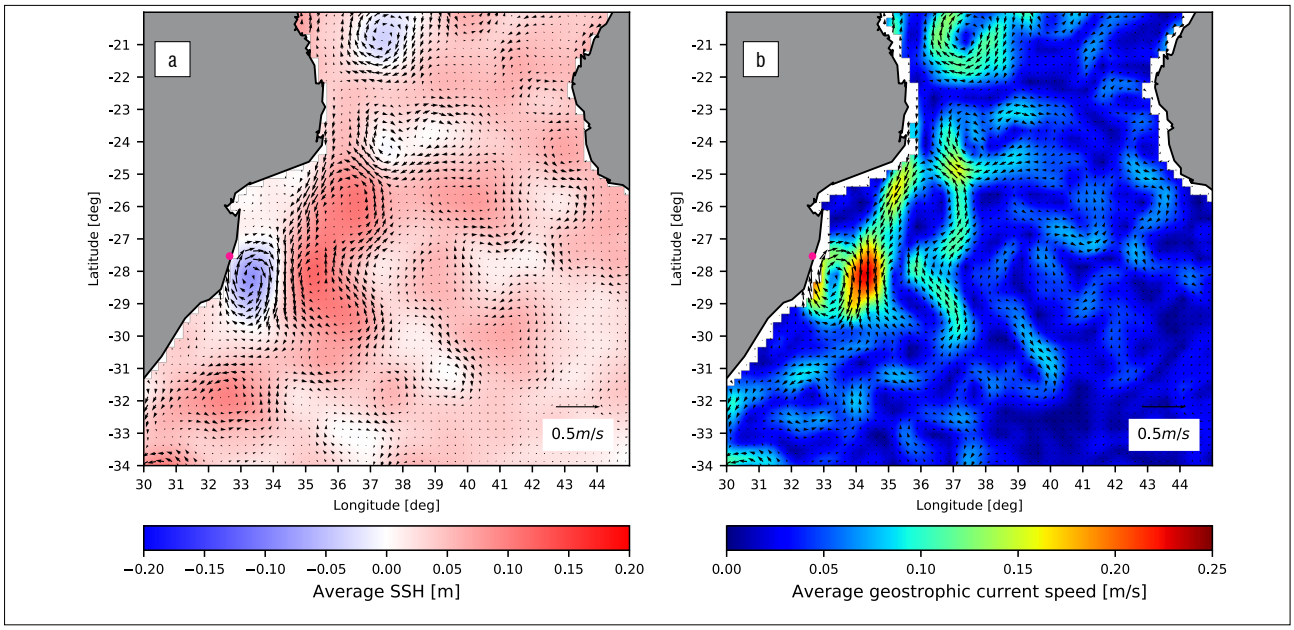


Figure 5: (a) Conditionally averaged sea surface height (SSH) field at the peak of the temperature anomalies using a -2 °C threshold. (b) Conditionally averaged geostrophic current field at the peak of the temperature anomalies using a -2 °C threshold. Geostrophic currents have been overlaid as a quiver plot and are scaled by speed.

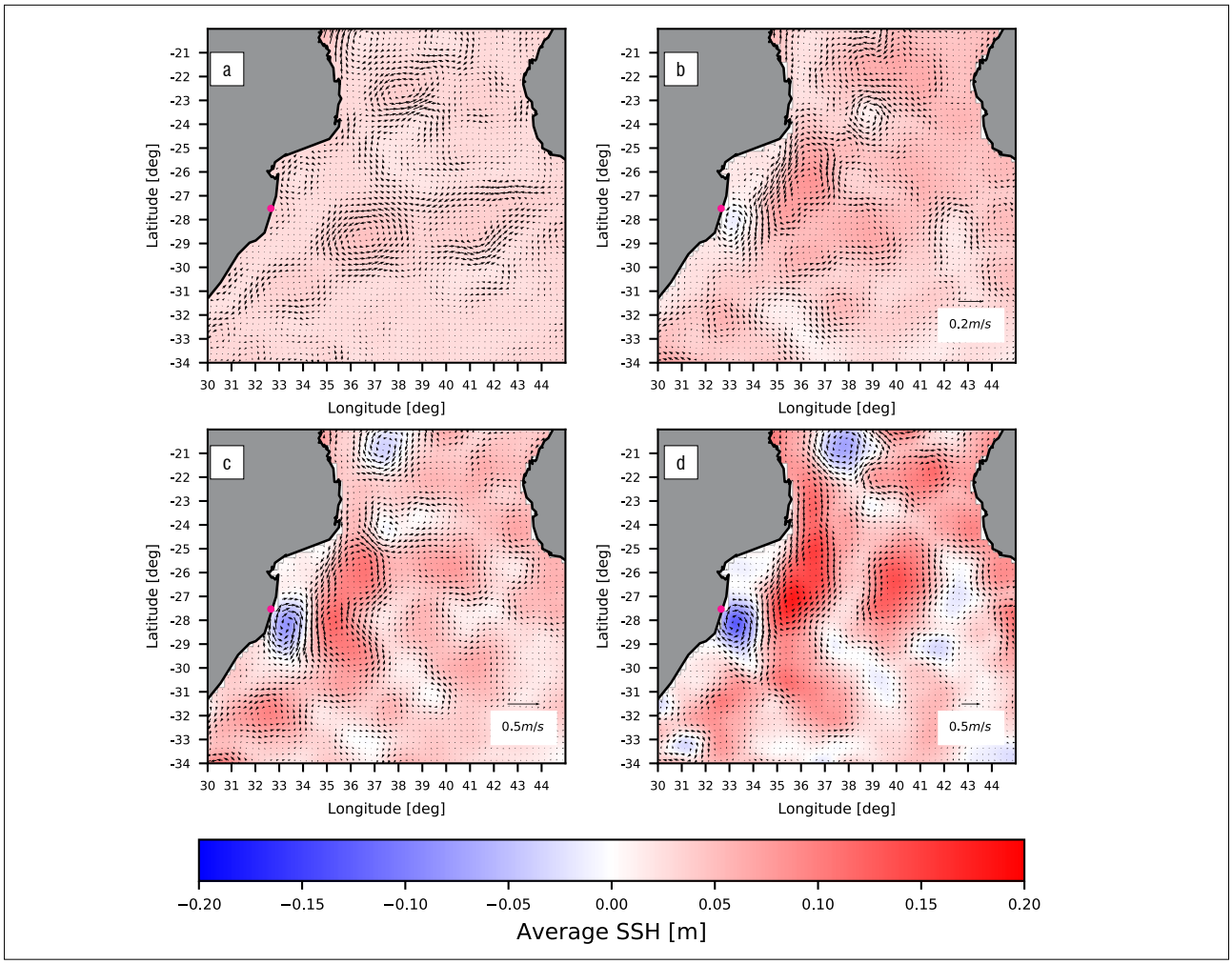


Figure 6: (a) Unconditionally averaged sea surface height (SSH) field. (b) Conditionally averaged SSH field at the peak of the temperature anomalies using a -1 °C threshold. (c) Conditionally averaged SSH field at the peak of the temperature anomalies using a -2 °C threshold. (d) Conditionally averaged SSH field at the peak of the temperature anomalies using a -3 °C threshold. Geostrophic currents have been overlaid as a quiver plot and are scaled by speed.

Of the 63 temperature anomalies identified, 60% of the events occurred during summer (1 December – 28 February), 17% occurred during spring (1 September – 30 November), 13% in autumn (1 March – 31 May) and only 10% of the temperature anomalies events occurred during winter (1 June – 31 August).

Conditionally averaged altimetry data

Figure 5 presents the conditionally averaged SSH field and the conditionally averaged geostrophic current speeds at the peak of the 63 temperature anomalies. Both plots have been overlaid with the conditionally averaged geostrophic currents.

A negative SSH anomaly and corresponding cyclonic eddy can be identified just offshore of the Sodwana region. The cyclonic eddy is constrained by the African coastline to the west and a larger anti-cyclonic eddy structure to the east. The negative SSH anomaly has an amplitude of 0.15 m and a cross shore diameter of approximately 200 km.

Temperature anomaly thresholds of -1 °C, -2 °C and -3 °C were investigated to explore their effect on the conditionally averaged SSH fields. SSHs and geostrophic currents at the peak of the identified temperature anomalies for each threshold were conditionally averaged. There were 211 temperature anomalies identified using a threshold of -1 °C, 63 temperature anomalies identified using a threshold of -2 °C, and 20 temperature anomalies identified using a threshold of -3 °C. The Kolmogorov–Smirnov statistic for the -1 °C threshold was 0.085 – lower than the critical value of 0.095. Therefore, the conditional SSH data set for the -1 °C threshold was not significantly different from the unconditional SSH data set. However, the Kolmogorov–Smirnov statistic for the -2 °C threshold was 0.319 – higher than the critical value of 0.172, meaning that the -2 °C conditional SSH data set was significantly different from the unconditional SSH data set. The SSH data set using a threshold of -3 °C was also significantly different from the unconditional SSH data set with a Kolmogorov–Smirnov statistic of 0.434 and a critical value of 0.304.

Figure 6 presents the unconditionally averaged SSH field along with the conditionally averaged SSH fields using various thresholds for identification of the temperature anomaly ranging from -1 °C to -3 °C.

In Figure 6, no distinct SSH patterns can be identified from the unconditionally averaged SSH. A negative SSH anomaly with a small amplitude near the Sodwana region emerges from the conditional average when using a threshold of -1 °C; however, this anomaly is not significantly different from the unconditional average. A clear negative SSH anomaly emerges from the conditional average when using a threshold of -2 °C, which is significantly different from the unconditional average. This implies that temperature anomalies associated with a drop of temperature greater than -2 °C can be linked to hydrodynamic patterns that are statistically significantly different from the mean hydrodynamics around Sodwana. The conditional average using a threshold of -3 °C shows the same negative SSH anomaly and cyclonic eddy pattern identified using a threshold of -2 °C. For thresholds lower than -2 °C, the conditional average pattern does not change.

The time lag of the conditional averaging of the SSH and geostrophic currents was varied around the peak of the temperature anomalies to investigate the temporal evolution of the conditionally averaged hydrodynamic patterns. Figure 7 presents the conditionally averaged SSH and corresponding geostrophic current vectors with various time lags. The conditional averages have been presented in 5-day intervals.

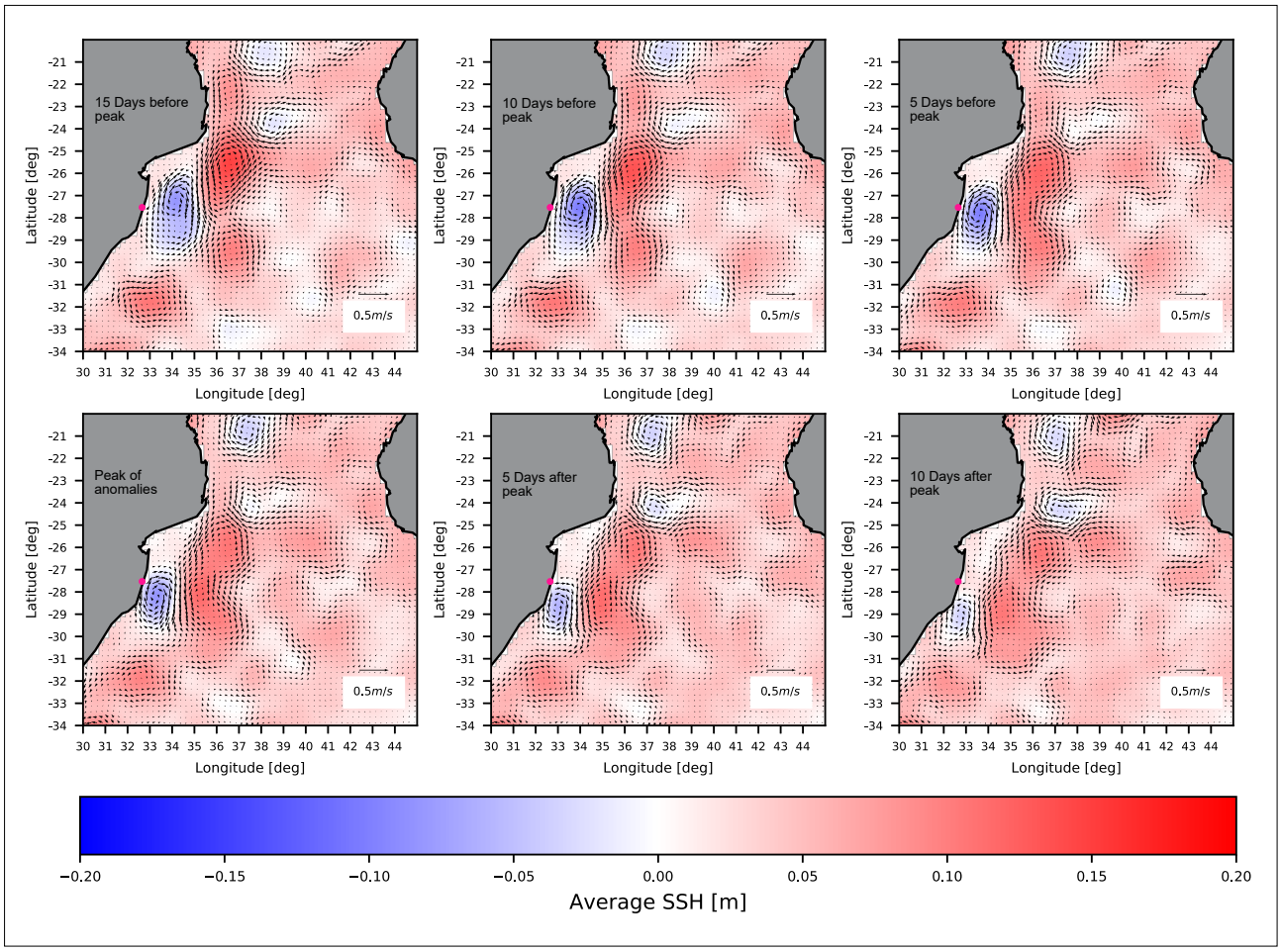


Figure 7: Sea surface height (SSH) and geostrophic current conditional average time lag between 15 days prior and 10 days post peak of the temperature anomalies. Geostrophic currents have been overlaid as a quiver plot and are scaled by speed.

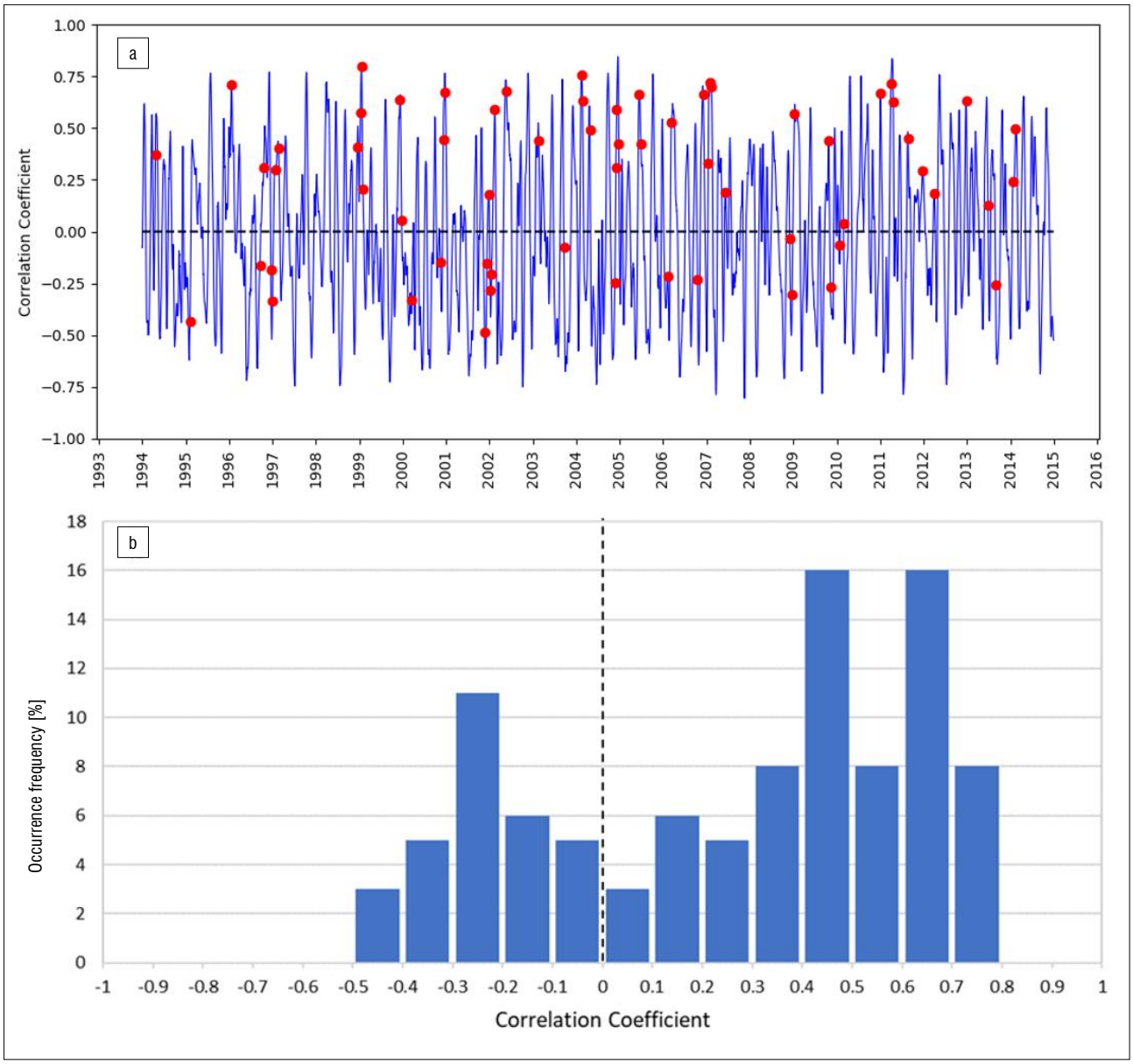


Figure 8: (a) Time series of cross-correlation coefficients. The temporal positions of the temperature anomalies are denoted by red dots. (b) Histogram of the cross-correlation coefficients of the sea surface height (SSH) at the peak of the individual temperature anomalies to the conditionally averaged SSH.

The conditional average 15 days prior to the peak of the temperature anomaly shows a well-defined anti-cyclonic eddy just south of the Delagoa Bight peninsula. A cyclonic eddy pattern can be observed to the northeast of Sodwana near the Delagoa Bight. The amplitude of the negative SSH anomaly is below 0.05 m and the cyclonic eddy structure is not yet well defined. The conditional averages from 10 to 5 days before the peak show that the anti-cyclonic and cyclonic eddy both propagate southward parallel to the African coastline. The amplitude of the negative SSH anomaly increases during this period. At the peak of the anomalies, the cyclonic eddy has propagated in a southwesterly direction and the centre of the eddy has moved past the Sodwana region. The southward propagation of the cyclonic eddy results in a net northward geostrophic current direction at Sodwana with a slight offshore component. At 10 days after the anomaly peak, both the anti-cyclonic and cyclonic eddies have propagated further south. The negative SSH anomaly has moved past the Sodwana region and has reduced in amplitude.

Altimetry data cross-correlation

Instantaneous altimetry SSH fields over the 21-year period were cross-correlated to the conditionally averaged SSH field within a 2° region around Sodwana. Figure 8 presents a time series of the cross-correlation coefficients over the 21 years. Also presented in Figure 8 are the temporal positions of the temperature anomalies denoted by red dots. This was done to show the correlation of the SSH of each individual temperature anomaly to the conditional averaged SSH. Figure 8 also presents a histogram of the temperature anomalies correlation coefficient values.

The correlation coefficients of the individual temperature anomalies range from -0.48 to 0.80. A negative SSH anomaly and cyclonic eddy offshore of the Sodwana region was observed for temperature anomalies with a positive correlation coefficient, which accounts for 67% of events identified. This means that 33% of the temperature anomalies were observed without the presence of a cyclonic eddy offshore of Sodwana. There are also many instances in which cyclonic eddies passed the Sodwana region resulting in a high correlation coefficient; however no

short-term temperature anomalies were evident. Only 15% of cyclonic eddies that passed the Sodwana region were associated with short-term temperature anomalies.

The temperature anomalies associated with the presence of a cyclonic eddy occur primarily in summer and account for 66% of the positively correlated temperature anomalies. Autumn accounts for 16% of the positively correlated temperature anomalies, winter accounts for 11% and spring accounts for only 7% of temperature anomalies associated with offshore cyclonic eddies. The temperature anomalies that were observed without the presence of an offshore cyclonic eddy occurred predominantly in summer and spring, with a 48% occurrence in summer and a 42% occurrence in spring. Winter and autumn account for only 5% of the temperature anomalies that were observed without the presence of a cyclonic eddy structure offshore of Sodwana.

The SSH and geostrophic currents at the peak of the temperature anomalies with positive and negative cross-correlation coefficients were then conditionally averaged separately. Figure 9 presents the separate conditional averages of the SSH at the peak of the temperature anomalies, with positive and negative cross-correlation coefficients.

The conditional average of the temperature anomalies with positive cross-correlation coefficients shows a similar pattern to the original SSH conditional average (see Figure 5); however, the anti-cyclonic eddy structure is stronger and more defined than in the original SSH conditional average. The conditional average of the temperature anomalies with negative cross-correlation coefficients shows a cyclonic and anticyclonic eddy pair situated approximately 200 km offshore of the Sodwana region. The cyclonic eddy structure would not have had any interaction with the Sodwana shelf due to its significant distance offshore of Sodwana. In the Sodwana region there is no negative or positive SSH anomaly pattern.

Cyclonic eddy tracking

The global eddy trajectory data[20] were used to identify the cyclonic eddies offshore of the Sodwana region during the temperature anomaly events. The purpose was to investigate the origin and evolution of the cyclonic eddies associated with the temperature anomalies at Sodwana. Figure 10 presents the 26 eddy trajectories that were identified passing Sodwana at the peak of the temperature anomalies. The seasonal temporal evolution of the cyclonic eddy trajectories has been included in the plot.

The trajectories indicate that the cyclonic eddies originate on the southwestern edge of Madagascar. The eddies then migrate westwards until they interact with the African coastline and subsequently migrate southward parallel to the African coastline. Occasionally the cyclonic eddy structures form in the Mozambique Channel and propagate southwards until they interact with the shelf south of the Delagoa Bight.

Discussion

The temperature measurements on Nine-Mile Reef, with the tidal and seasonal variance removed, show that temperature anomalies occur on average three times per year at Sodwana, and mainly in summer. This finding suggests that the frequency of anomalies provide potential relief from bleaching during the summer months. The temperature anomalies appear to regulate the temperature during the extended periods of elevated temperatures in summer and potentially improve the coral's resilience to bleaching.[14] Celliers and Schleyer[13] observed bleaching of the Sodwana reefs during the summer months of 1998 and 2000 when temperatures ranged between 27.5 °C and 28.4 °C. The temperature bleaching threshold of 27.5 °C was exceeded 18% of the time during the summer months over the 21 years. However, the bleaching observed by Celliers and Schleyer[13] coincided with a period during which no temperature anomalies were observed in the measured data (see Figure 4). This suggests that these anomalies could be important in mitigating bleaching at Sodwana, as noted by Riegl and Piller[11].

The conditional average of the altimetry data at the peak of the temperature anomalies indicates the presence of a negative SSH anomaly and corresponding cyclonic eddy offshore of the Sodwana region. The time-lagged conditional averages suggest the cyclonic eddy is present near the Sodwana region for a significant time during the anomaly period (see Figure 6). At the beginning of the time lag, the eddy is situated to the north of the Sodwana region and migrates southward at a rate of approximately 15 km per day. The cyclonic eddy structure also migrates shoreward towards the shelf over the duration of the anomaly period. After the peak of the anomaly, the cyclonic eddy continues to migrate southward and begins to dissipate.

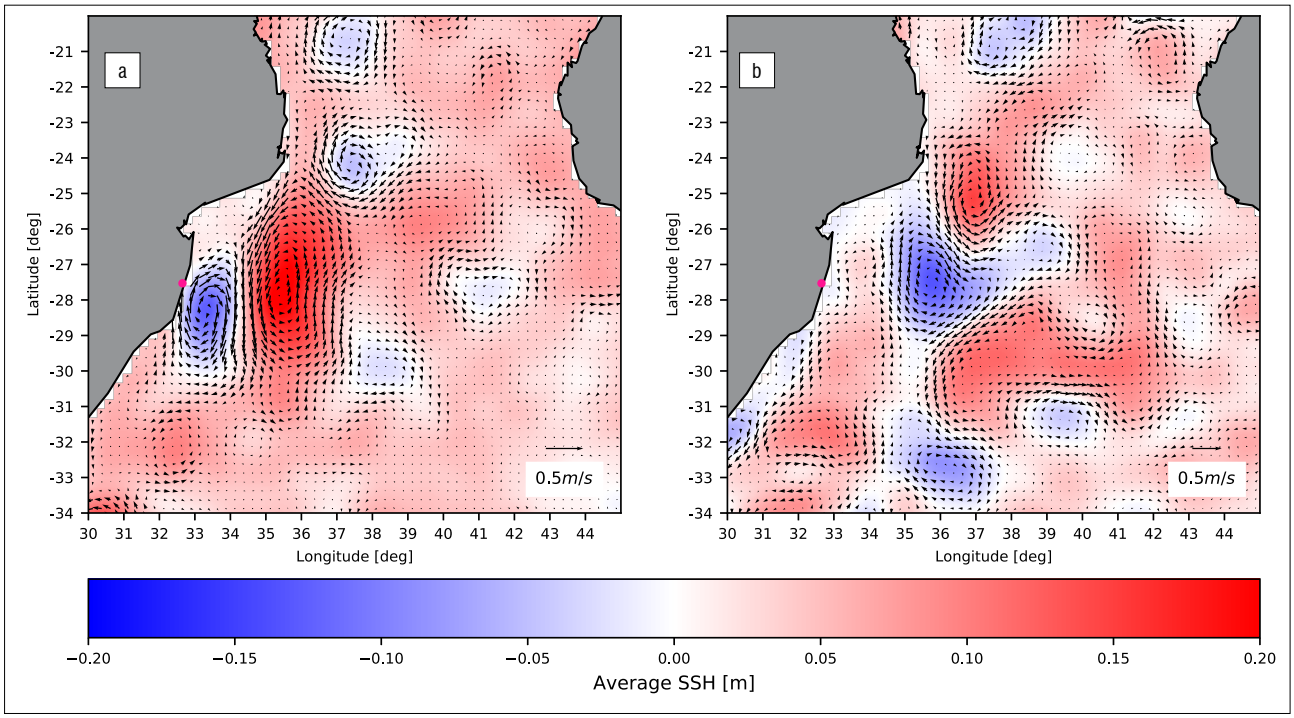


Figure 9: (a) Conditionally averaged sea surface height (SSH) field at the peak of the temperature anomalies with positive cross-correlation coefficients. (b) Conditionally averaged SSH field at the peak of the temperature anomalies with negative cross-correlation coefficients. Geostrophic currents have been overlaid as a quiver plot and are scaled by speed.

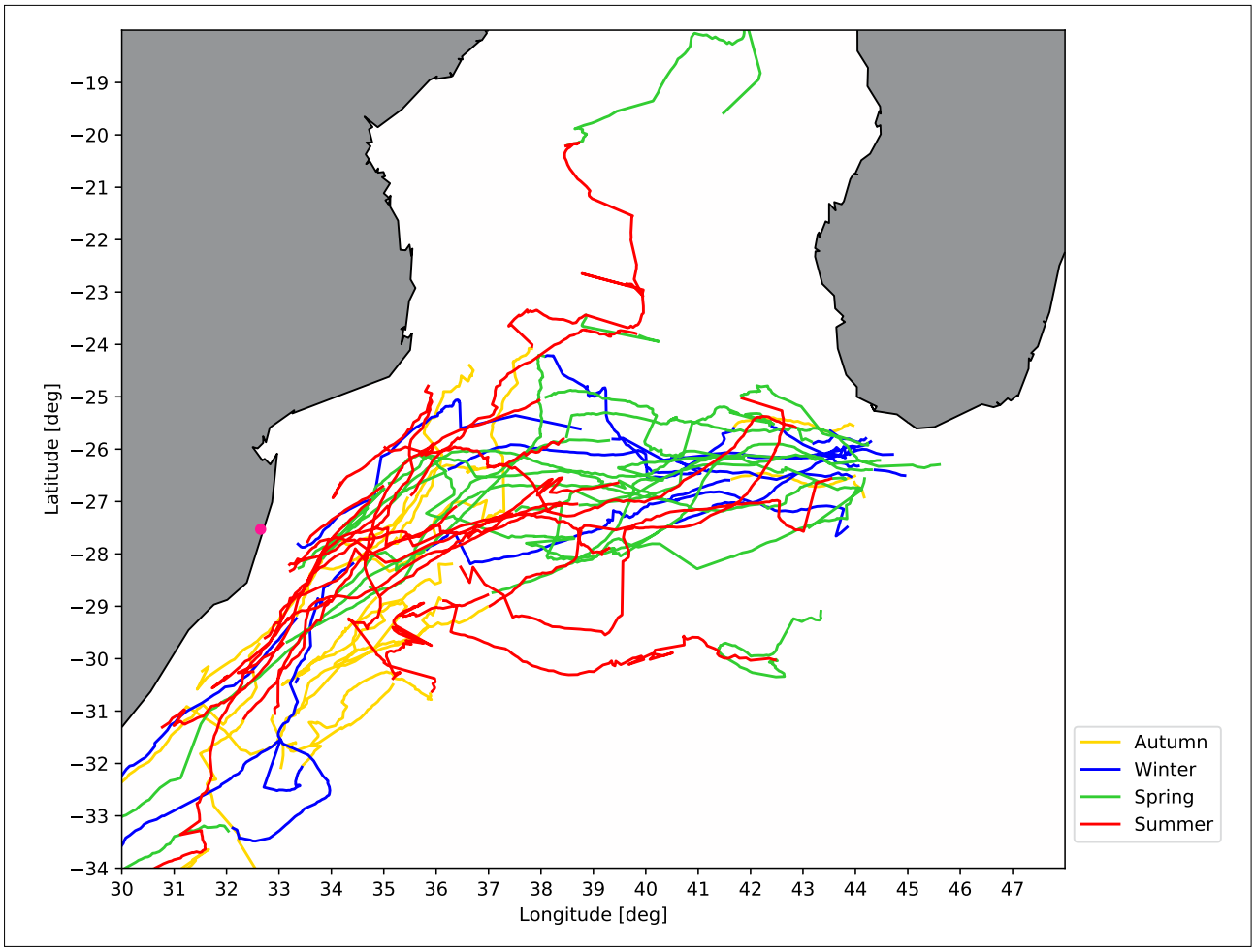


Figure 10: Cyclonic eddy trajectories that passed the Sodwana region during the temperature anomalies.

Roberts et al.[15] and Morris et al.[16] suggest that the presence of this eddy structure is associated with upwelling at the shelf edge and results in the temperature anomalies observed at Sodwana. They suggest that the interaction between the cyclonic eddy and the leading edge of corresponding anti-cyclonic eddy pull water off the shelf, which is replaced with cold water from deeper regions. This observation was based on a small subset of temperature anomalies that were investigated using altimetry data; however, the details of this mechanism remain unclear.

Another possible driving mechanism might be the reversal of the mean current along the shelf as the cyclonic eddies migrate past Sodwana. It is possible that this flow reversal interacts with the complex shelf bathymetry, advecting colder water up the canyons and onto the shelf. The interaction between the regional hydrodynamics and the Sodwana shelf and canyons has yet to be investigated and requires further research to explain this interaction and the resulting effects on temperature on the reef.

The cyclonic eddy trajectory results show that the cyclonic eddies associated with the negative temperature anomalies form in the lee of the southern tip of Madagascar. Quartly and Srokosz[30] observed that the majority of cyclonic eddies that interact with the South African coastline form in this region. Halo et al.[31] suggest that the formation of these eddies is driven by barotropic and baroclinic instabilities of the southeast Madagascar Current as it passes the southern tip of Madagascar. As these eddies migrate across the Mozambican channel, upwelling occurs in the core of the mesoscale cyclonic eddies bringing cooler deep water to the subsurface layers.[31] It is possible that when these eddies interact with the shelf near the Sodwana region, the cooler upwelled core may influence the temperature on the shelf.

The cross-correlation (see Figure 8) showed that the SSH and geostrophic currents at the peak of the anomalies are positively correlated with the conditional average for 67% of the temperature anomalies. This suggests that the SSH and geostrophic current patterns associated with 33% of the temperature anomalies at Sodwana differ significantly from the conditional average of all the temperature anomalies. A conditional average of the temperature anomalies with negative cross-correlation coefficients shows there is still a cyclonic and anti-cyclonic eddy pair at the peak of these anomalies. However, the cyclonic and anti-cyclonic eddy pair is situated approximately 200 km offshore of Sodwana near the Delagoa Bight region and does not interact with the shelf near Sodwana. The anti-cyclonic eddy structure is also situated further north of the cyclonic eddy as opposed to east of the cyclonic eddy as seen in the original conditional average. This suggests that a mechanism different from the eddy–shelf interaction is driving the temperature anomalies at Sodwana for these negatively correlated temperature anomalies. The cross-correlation results also show that there are many instances in which cyclonic eddies pass the Sodwana region, resulting in a high positive correlation coefficient; however, no short-term temperature anomalies were evident. This suggests that an offshore cyclonic eddy interacting with the shelf is not the sole driving mechanism of the temperature anomalies. Investigating these alternative mechanisms in detail requires further research.

Wind driven upwelling could be an alternative potential driver of the temperature anomalies in the absence of a cyclonic eddy offshore of Sodwana. Roberts et al.[12] suggests that northerly winds could

induce upwelling at Sodwana resulting in the cold water temperature anomalies. It was also suggested that wind driven upwelling could occur simultaneously with cyclonic eddy temperature anomalies leading to more intense cold water events.[12] Wind driven upwelling has not yet been investigated in detail and requires further research.

The results show that the temperature signal at a tidal frequency is well correlated to the tidal water levels. The temperature decreases with the incoming tide and increases with the outgoing tide. The results indicate that this temperature fluctuation is amplified over the duration of the temperature anomalies. It is possible that Sodwana experiences upwelling due to internal wave breaking on the shelf at the tidal frequency. This could drive cross-shelf transport of nutrients and cold water from deeper areas.[24,25] Masunaga et al.[25] suggests that the breaking of internal waves is dependent on the shelf slope and internal wave steepness. The complex shelf slope bathymetry around Sodwana may influence internal wave breaking and resultant upwelling in the Sodwana region; however, it has yet to be investigated in any detail. This driving mechanism of cold water onto the shelf is not directly linked to the presence of an offshore cyclonic eddy and may explain the temperature anomalies that occur without the presence of a cyclonic eddy.

Conclusion

We delineated hydrodynamic features associated with the occurrence of the cold water temperature anomalies around Sodwana. These temperature anomalies occur at Sodwana on average three times per year and predominantly during the summer months. The temperature anomalies can be associated with a negative SSH anomaly pattern and cyclonic eddy just offshore of the Sodwana region. The negative SSH anomaly and cyclonic eddy migrates southward past Sodwana at a rate of approximately 15 km per day and begins to dissipate after the peak of the temperature anomaly. It was also found that 33% of the temperature anomalies were observed without the presence of a cyclonic eddy offshore of Sodwana. This suggests that the presence of a cyclonic eddy offshore of Sodwana is not the sole driving mechanism of the temperature anomalies.

The regional hydrodynamic patterns associated with negative SSH anomalies and cyclonic eddies present possible driving mechanisms of cold water temperature fluctuations. However, we did not investigate the physics of each driving mechanism and the extent to which they drive the cold water temperature anomalies at Sodwana. Future research can use the results of this study to explore potential alternative drivers of temperature anomalies, for example, wind driven upwelling or upwelling due to internal waves breaking on the shelf at the tidal frequency.

It is recommended that additional in-situ data collection is required to improve future research on the driving mechanisms of the cold water temperature anomalies. The recommended in-situ data should include long-term current velocity and temperature measurements around the Sodwana reef system and submarine canyons. It is also recommended that numerical modelling of the flows around Sodwana shelf and reef is required to fully explain the driving mechanisms of the cold water temperature anomalies.

Acknowledgements

We thank the Oceanographic Research Institute and SAAMBR for access to the Nine Mile Reef temperature data. C.W. thanks the eThekwini Chair in Civil Engineering at UKZN for funding the research and PhD scholarship. D.S. is grateful to the eThekwini Municipality for sponsoring a Chair in Civil Engineering at UKZN. J.P. is supported by the South African National Research Foundation (UID: 127361).

Competing interests

We have no competing interests to declare.

Authors' contributions

C.W.: Conceptualisation, methodology, data collection and analysis, data interpretation, data curation, writing – initial draft, writing – revisions. J.P.: Conceptualisation, methodology, data interpretation, internal writing review, student supervision, project management. D.S.: Conceptualisation, methodology, data interpretation, internal writing review, student supervision, project leadership, project management, funding acquisition.

References

1. Souter DW, Lindén O. The health and future of coral reef systems. Ocean Coast Manag. 2000;43(8):657–688. https://doi.org/10.1016/S0964-5691(00)00053-3
2. Veron JEN. Corals of Australia and the Indo-Pacific. North Ryde, NSW: Angus & Robertson; 1986. https://doi.org/10.5860/choice.31-2102
3. Glynn PW. Coral reef bleaching: Ecological perspectives. Coral Reefs. 1993;12(1):1–17. https://doi.org/10.1007/bf00303779
4. Hughes T, Kerry J, Alvarez-Noriega M, Álvarez-Romero J, Anderson K, Baird A, et al. Global warming and recurrent mass bleaching of corals. Nature. 2017;543:373–377. https://doi.org/10.1038/nature21707
5. Brown BE. Coral bleaching: Causes and consequences. Coral Reefs. 1997;16(1):129–138. https://doi.org/10.1007/s003380050249
6. Dunn SR, Thomason JC, Le Tissier MDA, Bythell JC. Heat stress induces different forms of cell death in sea anemones and their endosymbiotic algae depending on temperature and duration. Cell Death Differ. 2004;11(11):1213–1222. https://doi.org/10.1038/sj.cdd.4401484
7. Glynn P. Coral reef bleaching: Facts, hypotheses and implications. Glob Change Biol. 2006;2:495–509. https://doi.org/10.1111/j.1365-2486.1996.tb00063.x
8. Wilkinson C. Status of coral reefs of the world: 2000. Glob Coral Reef Monit. 2001:18.
9. Lough J. 1997–98: Unprecedented thermal stress to coral reefs? Geophys Res Lett. 2000;27(23):3901–3904. https://doi.org/10.1029/2000GL011715
10. Buddemeier R. Is it time to give up? Bull Mar Sci. 2001;69:317–326.
11. Riegl B, Piller WE. Possible refugia for reefs in times of environmental stress. Int J Earth Sci. 2003;92(4):520–531. https://doi.org/10.1007/s00531-003-0328-9
12. Roberts M, Ternon J-F, Morris T. Interaction of dipole eddies with the western continental slope of the Mozambique Channel. Deep Sea Res Part II Top Stud Oceanogr. 2014;100:54–67. https://doi.org/10.1016/j.dsr2.2013.10.016
13. Celliers L, Schleyer M. Coral bleaching on high-latitude marginal reefs at Sodwana Bay, South Africa. Mar Pollut Bull. 2002;44:1380–1387. https://doi.org/10.1016/S0025-326X(02)00302-8
14. Safaie A, Silbiger N, Mcclanahan T, Pawlak G, Barshis D, Hench J, et al. High frequency temperature variability reduces the risk of coral bleaching. Nat Commun. 2018;9, Art. #1671. https://doi.org/10.1038/s41467-018-04074-2
15. Roberts M, Ribbink AJ, Morris T, Berg MA, Engelbrecht DC, Harding RT. Oceanographic environment of the Sodwana Bay coelacanths (*Latimeria chalumnae*), South Africa. S Afr J Sci. 2006;102(9):435–443.
16. Morris T, Lamont T, Roberts M. Effects of deep-sea eddies on the northern KwaZulu-Natal shelf, South Africa. Afr J Mar Sci. 2013;35:343–350. https://doi.org/10.2989/1814232X.2013.827991
17. Gruber N, Lachkar Z, Frenzel H, Marchesiello P, Münnich M, Mcwilliams J, et al. Eddy-induced reduction of biological production in eastern boundary upwelling systems. Nat Geosci. 2011;4:787–792. https://doi.org/10.1038/ngeo1273
18. Ramsay P. Marine geology of the Sodwana Bay shelf, southeast Africa. Mar Geol. 1994;120:225–247. https://doi.org/10.1016/0025-3227(94)90060-4
19. Schleyer M, Celliers L. Biodiversity on the marginal coral reefs of South Africa: What does the future hold? Zool Verh. 2003;345:387–400.
20. Faghmous JH, Frenger I, Yao Y, Warmka R, Lindell A, Kumar V. A daily global mesoscale ocean eddy dataset from satellite altimetry. Sci Data. 2015;2(1):150028. https://doi.org/10.1038/sdata.2015.28
21. Schlax M, Chelton D. The "growing method" of eddy identification and tracking in two and three dimensions. Corvallis, OR: College of Earth, Ocean and Atmospheric Sciences, Oregon State University; 2016.

22. Williams S, Petersen M, Bremer P-T, Hecht M, Pascucci V, Ahrens J, et al. Adaptive extraction and quantification of geophysical vortices. IEEE Trans Vis Comput Graph. 2011;17:2088–2095. https://doi.org/10.1109/TVCG.2011.162

23. Popiński W. Insight into the Fourier transform band pass filtering technique. Artif Satell. 2008;43:129–141. https://doi.org/10.2478/v10018-009-0012-9

24. Leichter J, Stewart H, Miller S. Episodic nutrient transport to Florida coral reefs. Limnol Oceanogr. 2003;48:1394–1407. https://doi.org/10.4319/lo.2003.48.4.1394

25. Masunaga E, Fringer O, Yamazaki H, Amakasu K. Strong turbulent mixing induced by internal bores interacting with internal tide-driven vertically sheared flow. Geophys Res Lett. 2016;43(5):2094–2101. https://doi.org/10.1002/2016GL067812

26. Adrian R. On the role of conditional averages in turbulence theory. In: Turbulence in liquids: Volume 1. Princeton, NJ: Princeton Science Press; 1977. p. 323–332.

27. Stretch D. Automated pattern eduction from turbulent flow diagnostics. Center for Turbulence Research, Stanford University: Annual Research Briefs. 1990:145–157.

28. Press W, Teukolsky P. Kolmogorov–Smirnov test for two-dimensional data. Comput Phys. 1988;2(4):74. https://doi.org/10.1063/1.4822753

29. Duda R, Hart P. Pattern classification and scene analysis. Libr Q. 1974;44(3):258–259. https://doi.org/10.1086/620282

30. Quartly G, Srokosz M. Eddies in the southern Mozambique Channel. Deep Sea Res Part II Top Stud Oceanogr. 2004;51:69–83. https://doi.org/10.1016/j.dsr2.2003.03.001

31. Halo I, Penven P, Backeberg B, Ansorge I, Shillington F, Roman R. Mesoscale eddy variability in the southern extension of the East Madagascar Current: Seasonal cycle, energy conversion terms, and eddy mean properties. J Geophys Res Ocean. 2014;119(10):7324–7356. https://doi.org/10.1002/2014JC009820

AUTHORS:
Coleen Vogel[1,2]
Mutizwa Mukute[3]
Kaera Coetzer[2]
Mzukisi Gwata[4]

AFFILIATIONS:
[1]School of Animal, Plant and Environmental Sciences, University of the Witwatersrand, Johannesburg, South Africa
[2]Global Change Institute, University of the Witwatersrand, Johannesburg, South Africa
[3]Environmental Learning Research Centre, Department of Education, Rhodes University, Makhanda, South Africa
[4]Environment, Infrastructure and Services Department, City of Johannesburg, Johannesburg, South Africa

CORRESPONDENCE TO:
Kaera Coetzer

EMAIL:
klcoetzer@gmail.com

DATES:
Received: 06 Feb. 2020
Revised: 25 Feb. 2021
Accepted: 28 Feb. 2021
Published: 29 July 2021

HOW TO CITE:
Vogel C, Mukute M, Coetzer K, Gwata M. Creating a climate of change in the City of Johannesburg: Co-learning to adapt to climate change. S Afr J Sci. 2021;117(9/10), Art. #7929. https://doi.org/10.17159/sajs.2021/7929

ARTICLE INCLUDES:
☒ Peer review
☐ Supplementary material

DATA AVAILABILITY:
☐ Open data set
☐ All data included
☐ On request from author(s)
☒ Not available
☐ Not applicable

EDITOR:
Yali Woyessa

KEYWORDS:
climate change adaptation planning, expansive learning, Cultural Historical Activity Theory (CHAT), co-learning, transdisciplinarity

FUNDING:
City of Johannesburg

Creating a climate of change in the City of Johannesburg: Co-learning to adapt to climate change

Climate change is one of the multiple stressors facing African cities; these cities are responding by developing climate change action plans including adaptation and mitigation policies. Effectively mainstreaming climate change in city plans and operations and moving from ambition to implementation is complex. Multi-actor engagement, transdisciplinary knowledge interactions, co-designing and sustained co-learning are often required in such planning and action contexts. In this paper, we trace and reflect on the process of developing an adaptation planning process for the City of Johannesburg, South Africa. Given shortcomings of the previous adaptation responses attempted in the City, specifically that of poor uptake, we trialled a more intentional and directly designed, formative and interventionist approach using Cultural Historical Activity Theory (CHAT). We reflect on what we as a research team and City officials learnt in this process. Our findings emphasise that exploring the local context remains critical in understanding and surfacing tensions with potential climate change responses. Failure to be mindful of such issues will likely result in mere compliance, and potentially, maladaptation. Contrary to experiences in other South African settings, rather than attempting to engage all actors simultaneously, our experience suggests that working with a core group initially, before expanding the circle of actors, is needed. These actors serve as mediators and pivotal actors for learning and change, and, with appropriate authority and passion, can drive, coalesce, and potentially re-enthuse waning interest from within. They leverage already existing trust relationships and strengthen participation throughout the process. Combined, these factors are critical for ensuring implementation and legacy.

Significance:

- Careful attention to a co-designed and emergent 'Theory of Change' can help both the process and design of engaged climate change research and help to reframe the climate action needed in urban contexts.
- The collaborative processes we applied increased awareness and engagement between officials around issues of climate change and, in particular, climate change adaptation.
- The lessons and opportunities gathered in the miniature expansive learning journey we trialled may be useful for others trying to embark on climate change adaptation journeys in cities in Africa and beyond.

Introduction

Climate change adaptation planning is a key process of national and local climate change policy in South Africa. The City of Johannesburg (CoJ), together with a South African public university, the Global Change Institute of the University of the Witwatersrand, embarked on a process to review previous climate change adaptation planning in the City and then co-reframe and co-design further adaptation action with the City. In this paper, we explore how an integrated urban climate change adaptation process can be developed using a theory-informed, iterative, transdisciplinary and multi-stakeholder process. We describe the approach we used to facilitate the co-development of a CoJ Climate Change Adaptation Framework (CCAF). This laid the foundations for the processes that were then expanded for the production of the Climate Action Plan (CAP) initiated in the City by C40 (https://www.c40.org/). We reveal the opportunities and challenges encountered and elaborate on how they were remediated and leveraged, detailing some of what has been learnt, with the aim to assist others engaged in similar climate change adaptation journeys in urban contexts.

Climate change adaptation planning and implementation action in cities has various methodological approaches, routes and pathways that can be taken.[1-3] These routes of action in a city are usually not linear progressions of a set of discrete engagements. Rather, they are made up of complex, messy interactions, that have several feedback and recursive loops of planning, action, 'back to the drawing board' journeys, rethinking and 'rejigging' and then embarking on another path of action. To enable this flexibility, we chose to ground the research in formative intervention, expansive learning and transdisciplinary approaches[4-6] with a strong emphasis on expansive learning[7].

Formative intervention usually involves collectively reframing a problematic situation and the development, application, refinement and integration of potential solutions in work and real-life issues. This generates potentially transformative learning and action.[8] Formative interventions have a practical interest to generate relevant solutions while meeting the rigour associated with academic research.[8,9] This is achieved by bringing practitioners, and content and process specialists together to co-develop, test, implement and refine solutions. We worked specifically with a formative intervention method called Change Laboratory, which was developed in the context of Cultural Historical Activity Theory (CHAT) to support expansive learning. We applied the principle of double stimulation to understand and reframe the challenges associated with climate change adaptation planning in the CoJ, in a manner that would account for their causes, which are often 'invisible'.

The City of Johannesburg's adaptation journey

Urban climate change adaptation in the City

Building robust climate change response is recognised as a critical component of strengthening the capacity of '… urban areas to face sudden as well as slow-moving risks'[10(p.39)]. The CoJ is a large metropolitan entity – the largest metropolitan municipality in South Africa, with an estimated population of around 5.74 million people.[11] Unemployment and youth unemployment in particular are high (32.7% and over 40%, respectively)[11] and contribute to a range of well-being challenges in the City. Details of the challenges (both economic and social amongst others) can be found in the City's Integrated Development Plan.[11] Rather than resorting to a technical review that focused on climate modelling and impacts and vulnerability assessments, we tried to situate ourselves *in the local context* of the City. In doing so, we were able to collaboratively explore some of the key multiple stresses which the CoJ faces (e.g. Bohle et al.[12], Cutter et al.[13,14]) and surface what it may take for the City and residents to better withstand future stresses that may be aggravated by climate variability and change.[15] Mindful of the various challenges the City faces, we grounded our research in the realities of the city actors (following Friend et al.[16]) and in the spaces that could potentially allow for transformative and integrated change (as in Mokwena[17]), particularly when linked to adapting to climate change.

Although the City had begun its climate change adaptation efforts in 2009, when a Climate Change Adaptation Plan (hereafter, '2009 CCA Plan') was developed, several challenges prevented it being mainstreamed into the City planning processes and everyday activities. Challenges included the lack of an integrated conceptual framework and approach which was needed to deal with the interlinked economic and socio-ecological issues of climate change impacts. The City also experienced difficulties integrating adaptation and mitigation, long-term and pressing short-term City needs, and encouraging ecosystem-based and hard infrastructure solutions into daily activities and long-term plans.[18] Most notable, however, were challenges associated with the aim to 'mainstream' all climate actions across all the City's departments and spatial areas. The absence of a clearly articulated, overarching climate change policy, in which the plan could be anchored, and the lack of clear financial and other incentives to support implementation are several issues that will be expanded on in this paper, which also further exacerbated the challenge of establishing an effective climate change adaptation plan in the City.

Such challenges are not, however, unique to the CoJ and have been explored, probed and investigated elsewhere in South Africa. In eThekwini (Durban) for example, the development of the Durban Climate Change Strategy and Climate Action Plan argues for the importance of a participatory and people-centred process – involving different departments, city residents, community groups, climate organisations, civil society groups and business.[3] Working in the City of Cape Town and reflecting on their engagement processes, Scott et al.[19(p.18)] cite Chu et al.[20] and Ziervogel and Parnell[21] who state that:

> *While adaptation outcomes are important, they need to be supported by adaptation as a process that builds capacity of different groups to adapt in different ways… These processes require time and expertise …* (emphasis added).

Complex governance and power relations[2,22], in various siloed departments in cities, also can work against integrative, systemic approaches for climate change adaptation planning[15,19,21,23-25].

In the CoJ, notwithstanding these challenges, various efforts over time have been made to begin planning for a broad climate change set of efforts, including a range of 'plans' linked to energy, water and disaster risk reduction, with few gaining any traction.[26] The CoJ, for a variety of reasons, has also been relying heavily on outside consultants to write their plans. Despite some value in this approach, the route we tried to pursue was a more engaged consultation process starting with people resident in departments within the City. More recently, the active engagement of C40 with an embedded person inside the City, builds on the work we helped initiate. This further influences the ways in which the City is now planning for climate change.[26] For example, the Environment and Infrastructure Services Department (EISD), C40, and various actors, have now crafted and designed a Climate Action Plan (hereafter CAP) which was officially launched in June 2021.

Methodological approach

Having been invited to engage to assist the City with adaptation planning, we proceeded to meet on several occasions with representatives from the EISD, the department responsible for climate change, to co-design a plan of action. The co-engagements between the City and the University research team were also enabled by a memorandum of understanding that had been signed between the City and the wider University. The combined research team of the City and University research team agreed on the research approach and ultimate goals of the adaptation planning and engagement strategy which extended over several months of the initial stage of the research. Ethical clearance was obtained from the University of the Witwatersrand's Human Research Ethics Committee to undertake this research (Certificate: H16/06/41) and a protocol obtained was updated to ensure ongoing research relevant to this work.

The research team consisted of five researchers, including an honour's student and an external facilitator who had nearly a decade of experience in working with CHAT-informed formative interventions and expansive learning in southern Africa. The team from the University was also supported by the Head of the EISD, her deputy and the lead for climate change adaptation in the City. Over the course of the research (a period of 3 years to date and still ongoing), the team was supported by interns (an embedded researcher) who also worked in the EISD. The interns were all previously trained by the University research team.

Using the CHAT formative intervention

The central focus in the engaged work that the collective team has been doing in the City is anchored in expansive learning involving the City and the research team. In trying to overcome the challenge of poor uptake and mainstreaming of climate change action in CoJ, this research trialled a more intentional, directly co-designed, formative and interventionist approach developed in CHAT. In CHAT-informed studies, the research team acts as a deliberate interventionist, creating an environment that allows practitioners to surface and critique their current activity based on its historical development and associated historically evolving contradictions. This provides methodological guidance for the creation of new practices through major redesign of the activity system.[7,27,28]

In CHAT, the basic unit of analysis that provides the minimal meaningful context for understanding human action is the *object-oriented activity system*, which is heterogeneous and multi-voiced.[29,30] An activity system is composed of different elements that interact and are driven by an object to produce something with societal value. The object is the issue that drives the activity system, giving it purpose.[31] Activity systems are dynamic in themselves and interact with other activity systems.[32] The elements of a contextualised second-generation activity system are illustrated in Figure 1.

Formative intervention results are emergent and not predetermined, similar to other transdisciplinary approaches.[4-6,31] The Change Laboratory method we used was preceded by the review of the 2009 CCA Plan and interviews with CoJ personnel to identify challenges in the Plan and in its implementation. Such approaches have been used to guide formative intervention research in different parts of the world and workplaces, including in southern Africa (e.g. Lotz-Sisitka et al.[33] and Mukute[34]).

Change laboratories are both a place and a process where practitioners, process specialists such as CHAT scholars, and content specialists such as climate scientists and adaptation experts, together with practitioners in the City (engineers, planners and other practitioners) meet to analyse historically emerging matters of concern around an activity or practice, identify contradictions and jointly reframe the concern and develop model solutions to address it.[9,35,36] Contradictions are viewed as 'historically accumulating structural tensions within and between activity systems'[37(p.137)] and manifest themselves in conflict of motives among participants in an activity or in interacting activities[36].

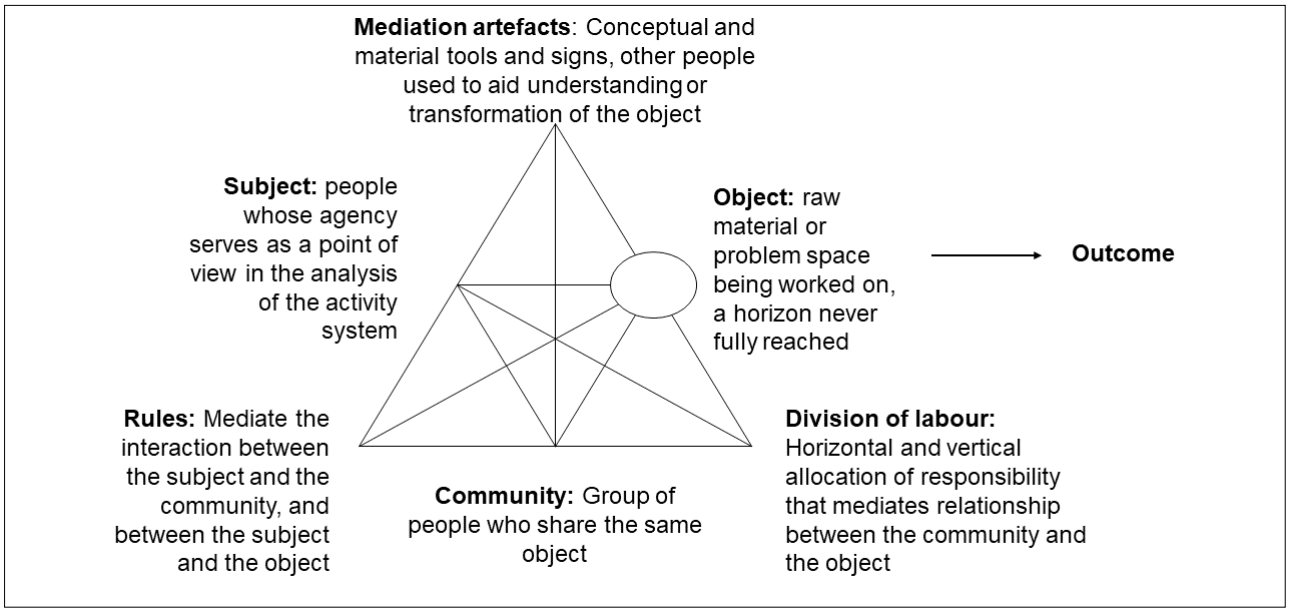


Figure 1: Elements of a contextualised second-generation activity system (adapted with permission from Engeström[7]).

The surfacing of different motives is intended to result in the development and implementation of solutions that will eventually transform work practices, relationships, objects and outcomes.[38] These transformations are achieved through applying the principle of double stimulation.[36] Double stimulation is a generative and intentional quest for change in which actors, such as organisations and people, deliberately break out of conflicting situations through joint problem-solving.[39] Double stimulation is an inherent and ongoing part of the process of expansive learning, which includes the following iteration:

- Questioning and analysing current activities and organisations, thereby identifying problematic matters of concern, which serve as the *first stimulus* for action;
- Identifying and using a concept that helps to reframe the problem and produce solution hypotheses is called a *second stimulus;*
- Identifying conflicts of motive and aspirations of multi-actors with a stake in the process and making choices and decisions that address the conflicts; and
- Taking volitional actions, i.e. making a deliberate decision to control one's behaviour and implementing the decision to break away from established constraints to implement the jointly agreed solutions.[39-42]

Change laboratories are relevant to tackling issues of climate change governance for at least three reasons:

1. They allow for the voices of different actors to be heard and considered in decision-making (at the very outset of any intervention), which is important from the perspective of democracy, agency and transformation.
2. Although individual viewpoints may remain different, the process of surfacing the variable understandings, motives and frames of reference has been found by Culwick et al.[43], in their work with CityLab forums, to lead to 'catalytic spaces of knowledge co-production'[43(p.13)]. In CHAT, the surfacing of these diverse understandings acts as sources of change, innovation and development of new solutions.[9,28]
3. Change laboratories also draw on the distributed knowledge, experience and interests of the different actors – practitioners and specialists – in problem analysis and the development of potential solutions.[38,43]
4. They emphasise transdisciplinarity of the climate change problem-space (i.e. across disciplines, sectors and regions). This is essential given the nature of climate change and climate change impacts, where objective scientific knowledge alone cannot account for the governance and/or social realities of 'what is happening on the ground'[44].

However, they do require considerable research and preparation before they are held; ongoing questioning and reflection between consecutive Change Laboratory sessions; time to develop mutual trust between the process researchers, practitioners and content specialists; time for collective learning, innovation and planning; time for implementing reviewing and refining solutions; time to change ways of doing things; and careful attention to issues of transformation learning and agency throughout the intervention research process.[34]

The City of Johannesburg as an activity system

We framed the minimum activity system as the CoJ (Figure 2). In this regard, CoJ can be viewed as an activity system that produces an assortment of goods and services for its residents, including water, housing, energy, human health and protection from the effects of climate change. CoJ also interacts with other activity systems such as the corporate sector, and the Gauteng provincial and South African national governments; activity systems thus do not exist in isolation.

The *object* carries the motive force of the CoJ activity system, in this instance, climate change and its impacts. The City's departments, that should operate as a collective but often do not, are responsible for addressing the impacts of climate change on the City's stakeholders through division of labour and the use of human, conceptual and material tools, and are guided by *rules* (policies regulations, conventions and standards). These rules mediate the relationship between the City and its stakeholders, determining how the City addresses climate change. Yet different departments have different perspectives on climate change and its impact (i.e. 'multi-voicedness'). Given this, the creation of a shared object between the City departments, i.e. where there is a jointly constructed understanding, is needed to support innovation in the climate action planning and climate change adaptation problem-space (after Kerosuo et al.[45]).

The City of Johannesburg's object can be viewed as a moving target for the City's adaptation response, and is unlikely to be reduced to and resolved in the short-term.[46] Instead, through the identification, development, refinement and application of *tools* of good and emergent climate change practice and policy, and supported through salient, credible and legitimate knowledge of both the urban and climate system[47,48], it may be transformed into the *outcome* of improved adaptive capacity to climate change over a long period[46].

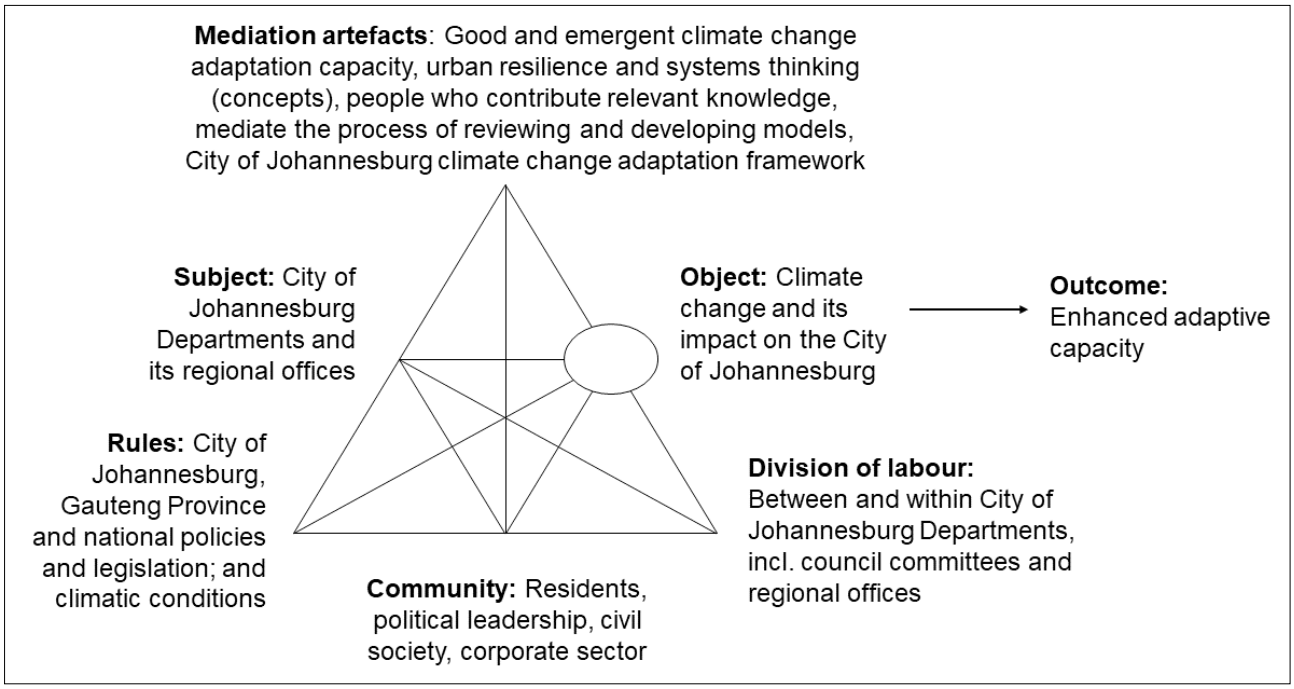


Figure 2: Activity system applied to the City of Johannesburg.

Sequence of the research steps

Our engagements with the CoJ were based on an expansive learning process and associated with four epistemic learning actions – *Questioning, Analysis, Modelling* and *Examining the new Model* (Figure 3, actions 5–7 still ongoing in the City) – which is typical of formative intervention research and a CHAT approach. The expansive learning cycle also, as outlined above, involves repeated cycles of applying the principle of double stimulation with first and second stimuli. The first stimulus is identified during the 'Questioning' and 'Analysis' Epistemic Actions 1 and 2, while the second stimulus is identified and used for the 'Modelling the solution' and 'Examining the Model', Epistemic Actions 3 and 4. These stimuli are subsequently used to identify, understand and refine the model solution in practice in the subsequent steps; and might result in the development and use of new solutions in an iterative process of review and analysis, solution development, implementation, review and refinement as the expansive learning process progresses.

Engeström, however, emphasises that the 'occurrence of a full-fledged expansive cycle is not common'[49(p.385)]. Instead, the expansive learning cycle more typically consists of smaller cycles of innovative learning (i.e. '*miniature expansive learning*'), which as an *iterative collective* can support overall organisational transformation in time. In this manner, and as we detail here, the expansive learning cycle (Figure 3) and its epistemic actions (Figure 4) provide the framework for analysing the smaller-scale miniature innovative learning processes we describe[49] and support by our continued relationship with the City.

During *Epistemic Action 1*, the '*Questioning*' phase, the research team conducted a document review of the original 2009 Climate Change Adaptation Plan (Table 1). The results of this review were presented to the CoJ as a first Change Laboratory ($n=27$), attended by representatives of EISD, Development Planning, Johannesburg Water, Environmental Health Department, Johannesburg Road Agency, Johannesburg Development Agency, Transport Department, and the Housing Department. This first Change Laboratory provided the forum for a discussion of the review process with the aim being to 'mirror' back, validate findings and further explore what the team had gathered about the City, planning processes and policies and co-design next steps in the adaptation planning process. After reflection from the Change Laboratory session, and keen to follow up on issues and 'matters of concern' that had been raised in the Change Lab session, attention then turned to undertaking smaller engagements.

A series of subsequent interviews with Development Planning, Spatial Planning and various members of EISD also enabled further engagement for probing identified issues further ($n=7$). These smaller meetings enabled the research team to engage more directly with City officials in a more intimate meeting setting and allowed us to better understand matters of concern that had been raised.

The research team also undertook a second document review of relevant CoJ documentation, including the Integrated Development Plans, Climate Change Strategic Framework and EISD's Climate Change Activation and Engagement Strategy.

At a second Change Laboratory with a group of an estimated $n\geq15$ participants *(Epistemic Action 2*: *'Presentation of findings, their validation and analysis'),* the results of the first Change Laboratory, interviews, and document review were confirmed, clustered and analysed for explanations by City departments. These first two learning actions ('Questioning' and 'Analysis') enabled co-determining why the original 2009 adaptation plan found no traction with City officials.

In *Epistemic Actions 3 and 4 ('Modelling'* and *'Examining the Model'*), a CoJ Climate Change Adaptation Framework (CCAF) was co-developed in various smaller Change Laboratory sessions – with an average of between 2 and 6 participants. Different departments responsible for the identified issues were tasked with reframing challenges and identifying potential short-term, medium-term and long-term adaptation actions and solutions to resolve them (*Epistemic Action 3: 'Modelling Solutions'*). These were incorporated and developed into a first-order draft of the CoJ CCAF. The development of this draft specifically sought to enhance and resource the process, conceptual and design limitations of the original 2009 CCA Plan.

The process of examining and improving the first-order draft CCAF (i.e. the 'model solution') took place at three levels, within and outside the CoJ (*Epistemic Action 4: Examining and Improving the Modelled Solution*). Firstly, we obtained and incorporated input from departments that had not been able to attend either of the Change Laboratories via interviews and consultations. The second level occurred through subsequent input from the interventionist researchers. Once the second-order draft CCAF was completed it was then presented to various technical and political committees in the City for wider endorsement.

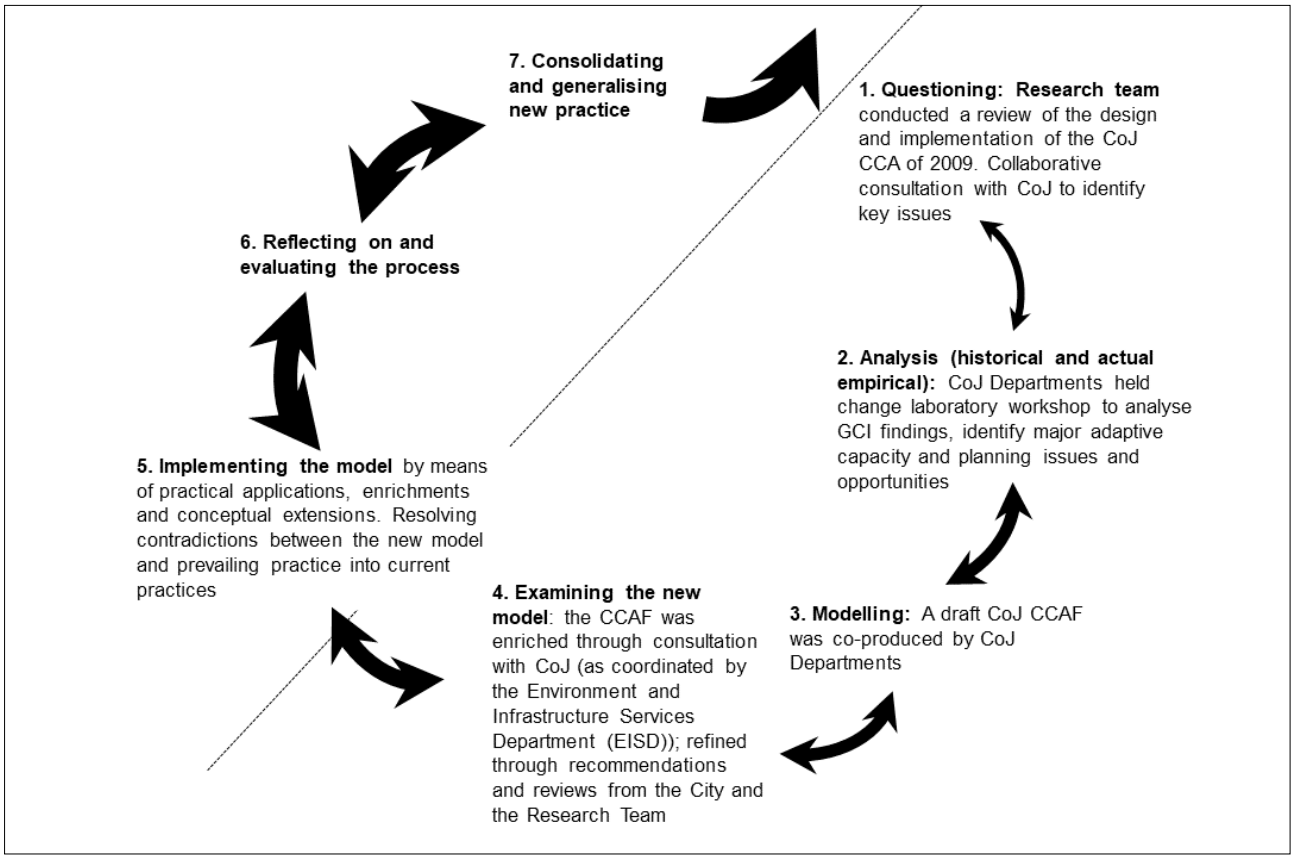


Figure 3: Expansive learning actions as applied in the City of Johannesburg's (CoJ) Climate Adaptation Framework (CCAF) development (adapted from Engeström and Sannino[9]). In the context of the results presented here, the development of the CCAF addressed epistemic learning actions 1–4. Actions 5–7 were still ongoing actvities at the time of writing this paper. Arrow thickness reflects scope for and level of participation in each action. CCA = Climate Change Adaptation Plan; GCI = Global Change Institute.

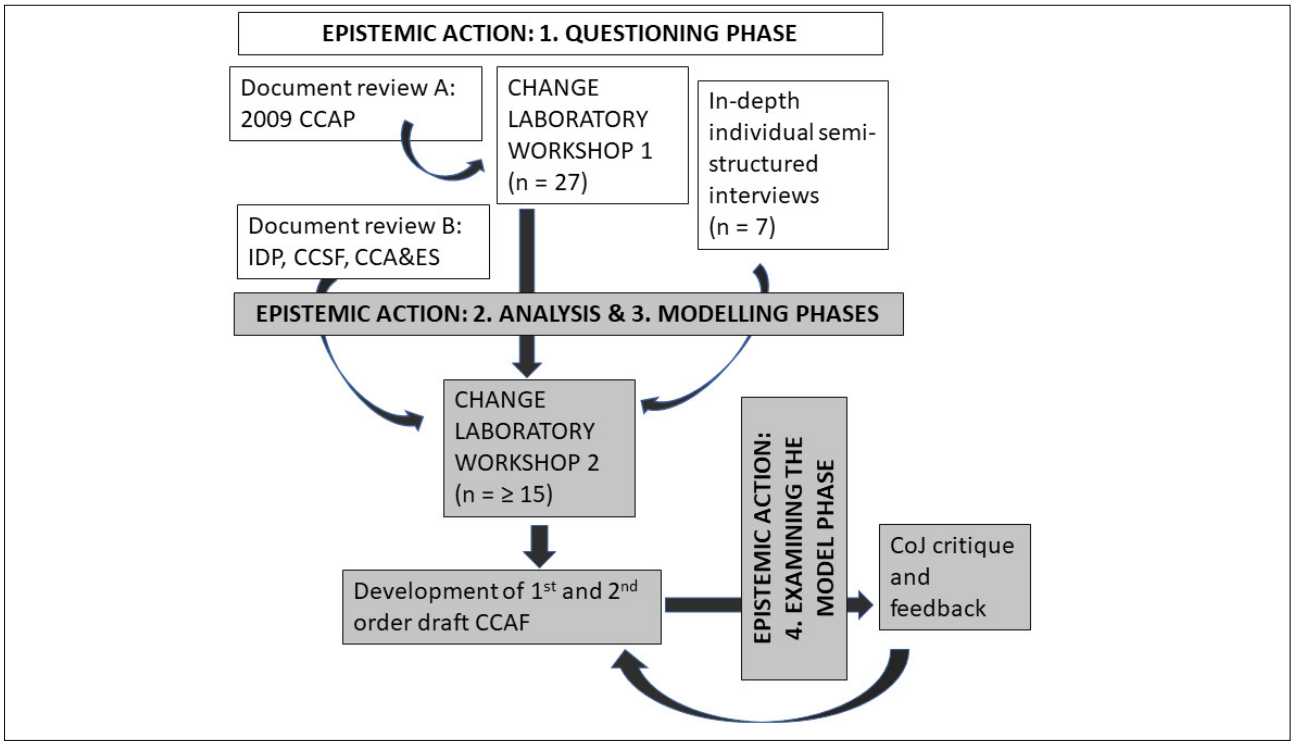


Figure 4: Research sequence of epistemic actions in the development of the City of Johannesburg's (CoJ) Climate Adaptation Framework (CCAF) including the action steps and methods used. The second Change Laboratory had more than 15 participants present (n=15–20), but participants arrived/left periodically due to other work-related obligations, with n=15 present for the full duration. CCAP = Climate Change Adaptation Plan; IDP = Integrated Development Plans; CCSF = Climate Change Strategic Framework; CCA&ES = EISD's Climate Change Activation and Engagement Strategy.

The document served before the Technical Sustainable Services Cluster meeting (a meeting of technical practitioners from EISD, City Power (Electricity Utility), Pikitup (Waste Management), Joburg Water, and Housing), and at the Mayoral and Section 79 Committees thereafter. The Mayoral Committee is chaired by the mayor and consists of all Members of the Mayoral Committee. Section 79 is a standing committee composed of various political parties and plays an oversight role, e.g. ensuring that departments use their budgets effectively. The Framework was well received at both Mayoral and Section 79 committees. These processes were key to follow and helped not only to ensure wider collective buy-in in the City, but also in expanding the learning in the City and enabling dialogues.

Results

A number of lessons and challenges have emerged through this climate change adaptation journey, several of which may have relevance and use for others embarking on similar processes in Cities. These are discussed below.

Epistemic Action 1: Questioning

The review process identified several matters of concern, surfacing tensions and contradictions between *short-term quick wins* and *long-term strategic* achievements.[18] Examples of such issues included *green issues* (biodiversity concerns) and *brown issues* (basic services for sanitation); *coping strategies* and *resilience building*; *abundance of water* (flood risk) and *water scarcity* (drought) risk management; *mitigation* and *adaptation*; national and provincial climate strategies and city climate strategies; and *hard (technocratic approaches*) and *soft green responses* (e.g. wetlands). In several cases, these 'opposites' were often voiced as tensions and contradictions. In trying to ensure a green, sustainable City, for example, water management processes anchored on wetland behaviour and a wetland's ability to transform/remove pollutants and increase infiltration and stormwater storage (e.g. Mander et al.[50] and Stefanakis[51]), were often seen to be at odds with the more traditional, 'hard' planning and technical city engineering design approaches (e.g. storm water drainage planning):

> *There are few integrated plans where all the activities of the City already underway (e.g., development of Sustainable Urban Drainage Systems in the Province and the City) are linked and connected to climate change adaptation planning in the City* (Interview respondent j).

In addition, a number of historical challenges and structurally embedded tensions[15] emerged. These included climate change matters of *immediate concern* (including the urgent need for service delivery in the form of safe water and sanitation and energy provision, particularly in informal settlements) and the need for identifying the vulnerable (which usually are also the result of driving factors that are the result of historical imbalances but also the result of the incapacity to deliver in a City with a range of other competing demands) that all remain persistent challenges.

Epistemic Action 2: Presentation of findings, their validation and analysis

Results from the various engagements with the City revealed and confirmed a number of emerging contradictions that again draw attention to issues of *process* being an essential part of designing planning in complex city contexts:

1. *Design:* the earlier 2009 CCA Plan was not linked to the Integrated Development Plan and Growth and Development Strategy, which occupy a higher place in the hierarchy of CoJ plans; and the councillors' and practitioners' Key Performance Indicators (KPIs), which matter in everyday work. The plan also did not have a learning, monitoring and evaluation component.
2. *Implementation:* The matters of concern that were identified in the present climate change processes described above were not dissimilar to those identified in 2009. CoJ's matters of concern, linked to climate change and climate variability, were confirmed as being water, waste, heat waves, food security, urbanisation and governance, particularly as related to planning and service delivery in the City. The expansive learning process, undertaken at least 7 years after the first Plan was drafted, furthermore validated and confirmed these issues as persistent climate change risks in the City.

These reconceptualised, interconnected (what some now refer to as intersectionality, e.g. in the case of gender and climate change[52]), problems defined the *first stimulus* in our study. The *second stimulus* resulted in the emergence of several conceptual tools enabling further ongoing interrogation of the 'matters of concern' and the development of possible planning actions. Key amongst these were the need to focus on adaptive capacity skills and development enhancement in the City, trying to build system-wide resilient city planning and where possible to focus on engendering complexity thinking. The ways in which performance is rewarded in the City (through siloed, key performance targets) will also remain a key barrier to effective mainstreaming in the City. Several City officials noted that, while they recognise the urgency for climate change action and implementation, the lack of incentives to do so and their narrow KPIs mitigate against a more transversal and systems' approach being possible.

A number of other persistent 'matters of concern' faced by City departments as part of their day-to-day challenges also emerged from the smaller group targeted interviews (Figure 5), and those seen as specific barriers to operationalising climate change considerations surfaced (Figure 6). Observations, made by several people we interviewed, noted the following persistent matters of concern:

- the need for more coherence and co-ordination across departments and entities – 'there is no inter collaboration on climate change' (Interview respondent j);
- the need to have job 'score cards' flexible enough to allow for issues such as climate change to be addressed with ease – 'how we are being corporatised prevents us taking up climate change as a key issue' (Interview respondent x);
- 'need the regions, ward committees and the departments in Braamfontein to better align' (Interview respondent x);
- 'the huge pressures of service delivery that can obscure other efforts that may be developed' (Interview respondent x); and finally,
- the need to make climate change an 'all citizens issue' – 'there is more outside of the City that lies in the hands of others e.g., business, civic engagement than just us officials inside the City' (Interview respondent x).

Through these various interactions, a shared climate change engagement that had not yet really emerged in the City began to emerge, with many colleagues acknowledging that the adaptation process which was set in motion through this work had enabled a joint understanding of 'each other's' departmental activities. In this respect, the Change Laboratory workshop processes and interviews offered a strong case for working with the expansive learning process. They allowed for the crafting of draft 'project management plans' whereby departments began to identify key concerns and to also begin to identify departments that could work together on such issues going forward. This modelling of possible solutions (i.e. departments identifying key priority actions with targeting of long-, medium- and short-term planning including identifying which City actors wanted to engage and work together) has begun to open up cross-sectoral and transversal learning processes that are still ongoing. As will be shown below, several of these engagements have enabled the compilation of the CAP where these early engagements had not gained policy traction.

Table 1: Sequence and description of epistemic actions in the conceptualisation and development of the City of Johannesburg's (CoJ) Climate Change Adaptation Framework (CCAF)

	Epistemic action	Action step	Description of epistemic action	Result of epistemic action
QUESTIONING	Epistemic Action 0: Preparatory – organise the practitioners, content and process specialists to work together	Engaged a competent facilitator with a sense of place (process specialist)	City of Johannesburg (CoJ) engaged a research institute at a South African public University to review the 2009 Climate Change Adaptation Plan (CCAP) and facilitate the development of a CCAF over a period of 1.5 years. The research institute was selected for its technical capacity on climate change and related responses, review and planning methodological strength and for being 'part and parcel' of the City.	Partnership between CoJ (public sector – as the practitioners) and research institute (research – as the content specialist on climate and climate change adaptation)
	Epistemic Action 1: Collaborative and inclusive identification of matters of concern using multiple methods	Conducted an inclusive and robust review of the 2009 CCAP	The review assessed how the plan was developed, implemented, resourced and monitored. It was further contextualised by assessing the CCAP in relation to the CoJ and national priorities, best practice and good examples of urban adaptation planning and implementation in South Africa and beyond. It utilised document analysis, literature review and interviews with CoJ representatives.	Identification of areas needing improvement
		Held Change Laboratory 1 to present and discuss findings of the 2009 CCAP review	CoJ organised a workshop in August 2016 at which the review was presented by the research institute. The Change Laboratory was attended by 27 professionals and practitioners from various CoJ regions and departments, namely: Environment and Infrastructure Services Department (EISD), Development Planning, Johannesburg Water, Environmental Health Department, Johannesburg Road Agency, Johannesburg Development Agency, Transport Department and the Housing Department.	Approval of the review Identification of lessons learnt and key issues Establishment of an inter-departmental Task Force to steer the process
		Conducted future-oriented in-depth interviews on key issues and potential climate change adaptation solutions	Seven in-depth interviews were conducted with representatives from CoJ departments over a period of 3 months in early 2017.	Identification of new and emerging climate change related matters of concern and opportunities
		Consulted recent studies on community awareness and concerns about climate change	Studies that were reviewed included a number of documents, but more recently: (i) the Integrated Development Plans (IDP) 2017, (ii) Update on the IDP process, (iii) Climate Change Strategic Framework, 2015 and (iv) the Climate Change: Activation & Engagement Strategy prepared for EISD	Identification of climate change awareness levels among urban communities and their climate change priority issues
ANALYSIS	Epistemic Action 2: Collaborative problem analysis, elaboration and prioritisation	Presented matters of concern which were jointly analysed and prioritised	A synthesis of matters of concern was presented and discussed at a second Change Laboratory held on 2 June 2017, which was attended by members of the Task Force/Steering Committee, regional managers and staff from CoJ departments. These matters were confirmed and analysed for explanations.	Joint validation and elaboration of identified matters of concern Collective prioritisation of matters of concern Identification of linkages between the concerns
MODELLING	Epistemic Action 3: Co-development of the CoJ CCAF	Developed response strategies and actions to address matters of concern	Interconnected issues were clustered together and departments responsible for their tackling and resolutions worked together to reframe the challenges and develop strategies to address them.	Identification of the focus of the CCAF Consensus on the conceptual framework(s) to work with Outline of strategies to respond to climate change impacts in CoJ
		Compiled the draft CoJ 2017 CCAF	The 2017 draft CCAF was developed by drawing on the contributions and outputs of all the preceding epistemic actions, which were based on inputs from some of the CoJ stakeholder groups and CoJ strategic documents.	Draft CCAF with direct input from CoJ departments and indirect input from other stakeholders
EXAMINING THE MODEL	Epistemic Action 4: Examining and improving the CoJ CCAF	Presented and discussed the draft CCAF with stakeholders for their input	The priorities and main strategies of the draft CCAF were disseminated in face-to-face meetings with stakeholders in July 2017. Stakeholders gave feedback, which was incorporated in the final draft 2017 CCAF and also included into the Climate Action Plan (CAP) C40 process.	Widening of stakeholder input and increased potential for relevance and co-ownership of the CCAF

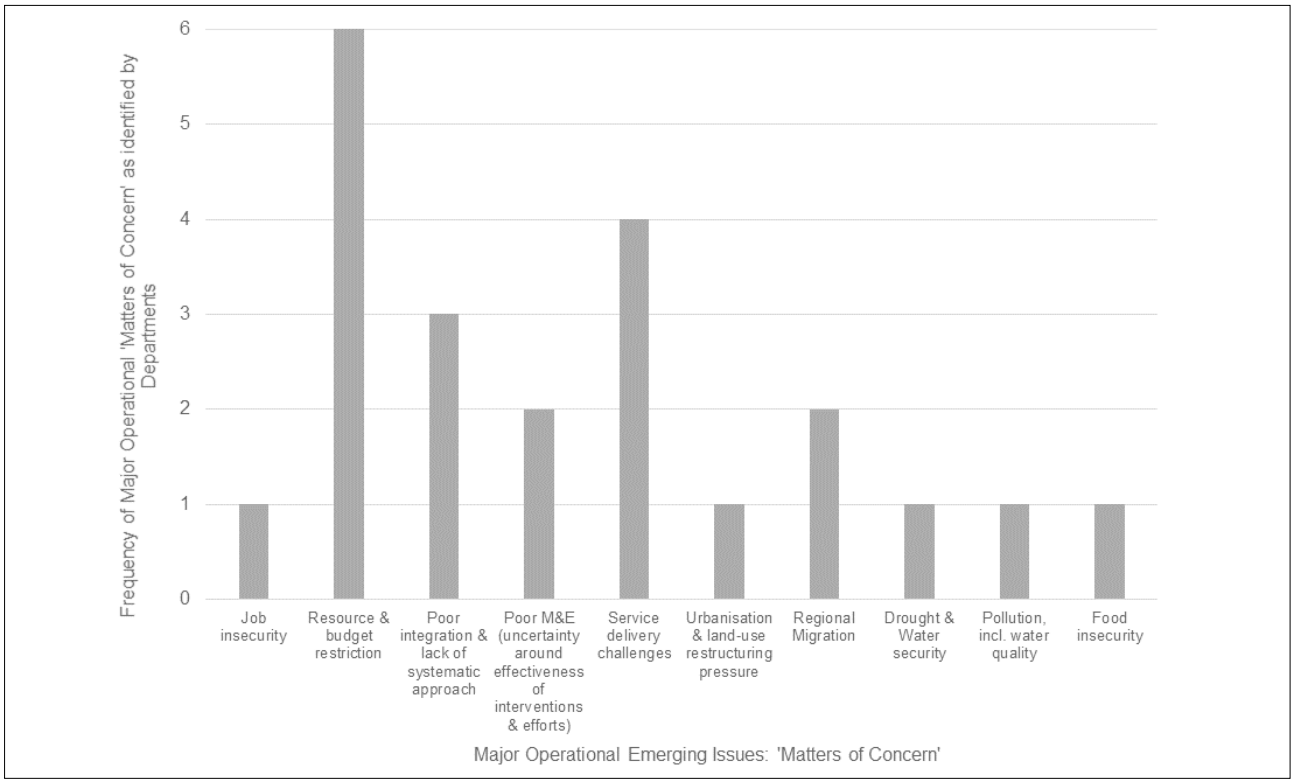


Figure 5: Emerging issues ('matters of concern') seen as major operational challenges in the City of Johannesburg's (CoJ) day-to-day activities. Operational concerns were thematically coded from individual one-on-one and small group targeted interviews conducted with representatives of different CoJ Departments (City Parks, Environment and Infrastructure Services Department, Development Planning, Disaster Management, Johannesburg Water, Environmental Health Department, Johannesburg Road Agency). Frequency is the aggregated count of themes that emerged from interview data across all Departments.

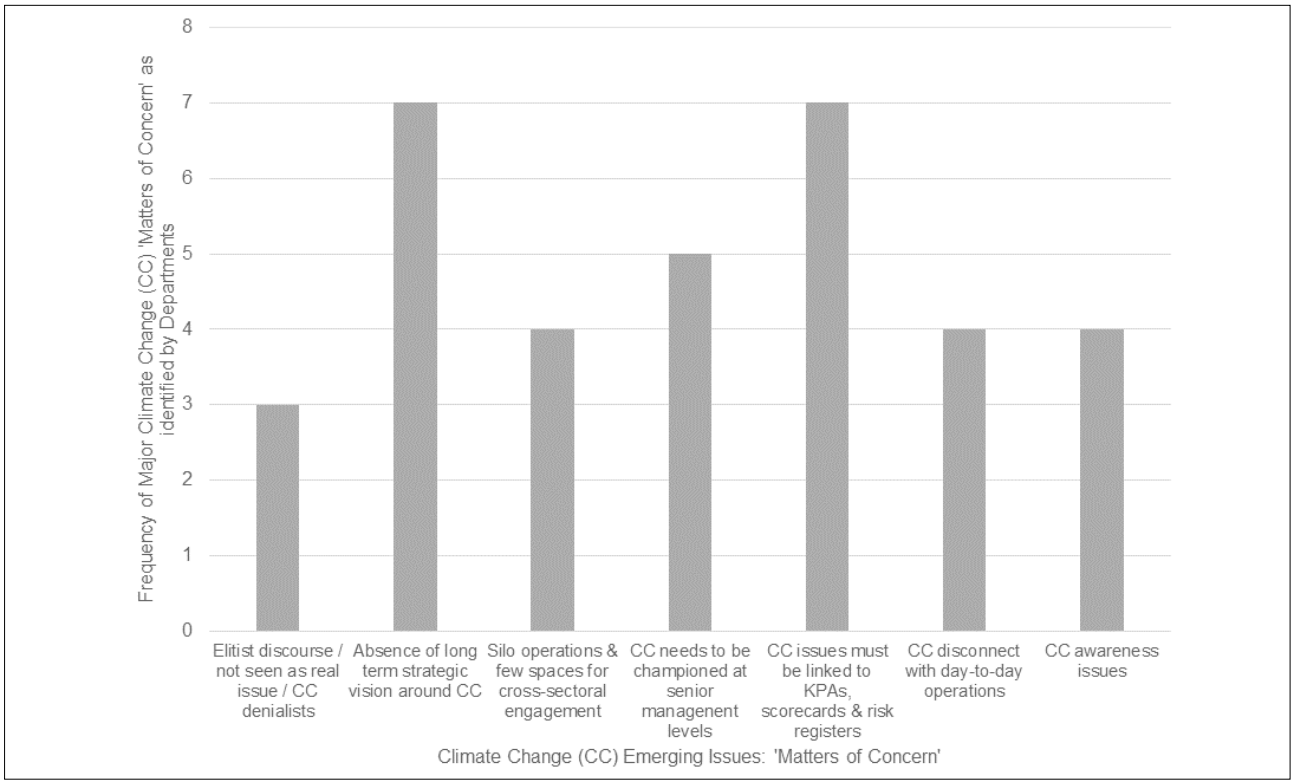


Figure 6: Emerging issues ('matters of concern') seen as potential barriers to operationalising climate change ('CC') in the City of Johannesburg's (CoJ) activities. Climate change related emerging Issues were thematically coded from individual one-on-one interviews conducted with representatives of different CoJ Departments (City Parks, Environment and Infrastructure Services Department, Development Planning, Disaster Management, Johannesburg Water, Environmental Health Department, Johannesburg Road Agency). Frequency is the aggregated count of themes that emerged from interview data across all departments.

Epistemic Actions 3 and 4: Modelling solutions and improving the 'model' solutions

The epistemic actions described above culminated in the production of the first draft CCAF. This was used as a document that could circulate through various forums in the City and receive comment and inputs, widening stakeholder input and increasing potential for relevance and co-ownership of climate change planning. In the final stages of this process of adaptation efforts in the CoJ, the City also embarked on the development of a CAP in collaboration with C40, that built on and included sections of the adaptation work described here, as well as climate change mitigation efforts.

The CAP, a mayoral agreement signed with other African City mayors all committed to a planning process and policy for climate change action, was published in March 2021. At the time of finalising this paper, the CAP had gone through the political process of approval (19 November 2020) including the approval of the draft CAP and the establishment of a Climate Action Forum going forward that will assist in reporting and the implementation of the CAP. The CAP will also be reviewed every 5 years. The Council still had to review the CAP at the time of finalising this paper (February 2021), but only a few months later, by 3 June 2021, it had been officially and publically launched.

The next steps, in which the research team has remained involved and actively engaged, include engagement with the youth and wider civic society so that climate change actions can be further designed and mainstreamed in the City and the 12 regions. The youth with whom the team has been engaging (details of which are being prepared for another publication) were able to have their inputs included directly in the CAP document – a major achievement for the youth cohorts engaged as well as the research team.

Discussion

In conducting a learning-oriented climate change adaptation process, we aimed to work in a very complex setting, both politically and institutionally, but also as honest and humble brokers of climate change action. Since the time of the commencement of this research (late 2015, early 2016) we have attempted to have a very mindful and mutual learning journey and relationship with City officials and practitioners. This relationship has included being mindful of the complex tensions that comprise the 'local context' in the City, being able to collectively identify the 'matters of concern' in the City, and then examine how these could be linked and actioned with climate change efforts. Rather than address these in an item-by-item fashion, we reflect more holistically and reflexively on what has been learnt through a process of expansive learning in the City.

The journey, as illustrated in this paper, is a long, and often frustrating one, with several moments of both an *outward* (more academic and bureaucratic learning in the City, i.e. learning about their procedures and policies and also our own academic theoretical learnings) and then an *inward, personal journey (*of self-questioning and growth), both for the research team (including the City personnel) and the actors in the wider 'outside' City we met and worked with.

Some of the 'outward' academic learnings that are emerging from this intervention research and are still ongoing are discussed below.

The main subjects of an activity system may serve as an important entry point of formative intervention. In complex activity systems such as urban Johannesburg, where there are many stakeholders including residents, councillors, the business community and civil society, it is not feasible to engage all the actors simultaneously. The City's original 2009 climate adaptation efforts lacked the participation of key stakeholder groups, including City departments, residents and the business community in issue identification. It was considered a notable limitation of the initial approach – perhaps as a result of the difficulties in engaging effectively with this diverse actor network.

However, our experience suggests that leverage can be achieved through working with a *core group of people first, i.e. key actors, before expanding the circle of actors*. By spending much time with the EISD and then meeting various department heads and others in the City, we began a trust-building journey that earned us some credibility in the City. Such an approach runs counter to other similar engagements where efforts were focussed on 'outside engagement' simultaneously with internal (e.g. Durban/eThekwini[53,54]). This departure from the other cases was enabled by very strong champions who were working inside the City and thus could help navigate more clearly the various forms of engagements, both internally and externally. The City now is planning subsequent, more varied actor engagements (e.g. expanding the learning to youth change agents and civic society in the regions).

Some of the subjects can also serve as mediating tools for learning and change, in the form of champions. The lead department on climate change, EISD, serves as the initial champion in the process of generating stakeholder buy-in and co-production of knowledge and potential climate change adaptation solutions. Having a lead department or individual that has the authority and the passion to drive, coalesce, and, potentially, re-enthuse waning interest when necessary *from within*, cannot be emphasised enough. When attempting to generate and consolidate new practice, the role of embedded champions becomes invaluable as they are able to leverage existing trust relationships and strengthen participation throughout the process.

Conflicts of motive among participating actors need to be carefully surfaced and contextualised. Cities operate in very siloed ways, with each department tied to specific deliverables. Engaging in dialogues and co-learning environments that are not sensitive to such contexts can easily mean disengagement with local actors at all scales and levels. The advantage of therefore spending time understanding the local context in which actors find themselves (e.g. in the expansive learning cycle; Figure 3), in understanding and surfacing tensions and contradictions cannot be underscored more heavily. Roberts[53(p.536)], in reflecting on developing Durban's Climate Strategy has similarly emphasised that embedding climate change considerations in the local context requires that it is 'framed within a broader social/environmental justice framework', and thereby, *internalised for local context and priorities.* This ensures the local development agenda, with its associated resource (human, time, financial) allocations, becomes more meaningfully associated with robust climate protection response; *actors need to see that their 'matters of concern' do indeed 'matter*'. Lack of doing so results in a lack of cogency in the process and can result in mere compliance (i.e. 'tick box' activities), which may result in maladaptation.

Applying the principle of double stimulation potentially enables generative learning processes. A number of tensions and contradictions surfaced at the very beginning of this process and are being probed further with other constellations of actors, e.g. civil society, mayoral executives, as the adaptation planning process continues. Issues that surfaced not only became objects of interest in themselves, but also informed the probing methodologies applied, such as double stimulation and conceptual tools on adaptation. However, *double stimulation is not a one-off engagement process but rather a sustained change process requiring patience and dedication.* Patience and considerable time are needed to develop potentially transforming and transformative climate change adaptation frameworks and plans.

Several contradictions, opportunities and 'inward' learnings also emerged in this work for the research team, including City actors who were part of the team. One of the main challenges that emerged is the challenge to one's own political value system and beliefs and what type of 'political context' can be considered for climate action planning going forward in the City. At the commencement of the project, the City's leadership changed from the African National Congress (ANC) to the Democratic Alliance led coalition government and more recently has changed back to the ANC. These changes in leadership did not only bring with them changes in political paradigms and leadership styles, but also the attention given to climate change (Figure 6) over time.

The political drivers of change in the CoJ, as with any City, are also very contested. As others have noted (e.g. Hetz[55]), the tasks of bolstering economic growth and yet also attending to socio-spatial disparities in

the City, are dilemmas for the City in both deciding future development trajectories and how the City responds to climate stresses:

> *In Johannesburg, the operational response space for climate adaptation is likely to be determined by the ways adaptation practices can be synergised with the pressing planning challenges of the urban divide, and how, accordingly, adaptation practices gain political legitimisation in reference to society's expectations about planning outcomes.*[55(p.1176)]

Another significant contradiction that the research team and the City is facing is that linked to the time horizons for planning and how climate change actions are currently being perceived and considered. Despite the focus on a resilient, sustainable and a livable city as expressed in the City's Growth and Development Strategy 2040, climate change '...responsiveness to planning is not a priority criterion'[55(p.1176)]. In this respect, officials repeatedly emphasised the absence of a long-term strategic vision tied to climate change as a significant obstacle for climate change appearing on project radars, with traditional employee performance measures reinforcing this (Figure 6).

One possible reason for the lack of an 'adaptation focus' and a focus that is more structurally orientated and transgressive[33] is that of time scale. Undertaking a more proactive approach, with a longer-term adaptation planning view, is often circumvented by a focus on short-termism, political cycles, and projects that produce highly visible outcomes such as the current focus on service delivery.[55] Similarly, work done by Polk[6] has indicated that tensions, obstacles and perceived delays in generative processes that typify co-designed research and practice, may cause already over-burdened participants to revert to familiar roles and 'ways of doing'. The research team are anticipating that now that the CAP has been launched and implementation has begun, new kinds of challenges are also going to arise and require the continuation of the expansive learning journey towards embedding the new CAP into the everyday practice of the wider CoJ. Patience and trust building are thus required in such a mutually engaged process and learning to ensure commitment to longer-term, slower-to-deliver outcomes.

The way officials are rewarded through their operational 'score cards' thus can also frustrate ongoing processes that are not suitable for the types of reflexive learning process we were collectively trying to share. The KPIs and siloed approaches in the City encourage a more compliance management style (e.g. ticking boxes) rather than a more open, experimental and social learning approach. The ways in which the City is administered, both bureaucratically and financially, present challenges. Finances are allocated, released and aligned according to capital budgets (e.g. through the Johannesburg Strategic Infrastructure Platform). Budget cycles also constrain longer, more generative processes that may be required for effective climate change adaptation 'learning', as cycles tend to be short and highly competitive. Consideration of how budgets are allocated and ways to support those with capacity constraints is '... an important step to specifically addressing environmental risks via infrastructure planning'[55(p.1177)], as for example, a gap of about '… 500 mill ZAR capital budget for the proposed storm water projects exists'[55(p.1178)]. These challenges are exacerbated by various issues related to the management and day-to-day running of departments and related activities (Figures 5 and 6).

Finally, more personal reflections and challenges that emerged in both the research team and some of the City officials, were coupled to the tension of wanting to profile climate change and yet also wanting to ensure that climate change actions were embedded in the mainstream without detracting from current, urgent developmental planning in the City. One is very mindful that a climate justice dimension is critical in all such climate change work and that *much* more engagement with citizens is needed. Related to climate justice, one constantly questions one's own position in such work – the concerns, for example, about inter-generational climate justice. Is the goal to move towards reducing the impacts of climate change now through the design of a climate change adaptation plan, or should the planning be more generative, that is what may be required *now* so that we can all 'live' with climate change in the future? Perhaps the focus should instead be on immediate 'matters of concern' such as water, food, shelter and energy security in the now, hoping that some spin-offs may emerge in the future for longer-term climate change adaptation? City officials, thus repeatedly, noted existing service delivery challenges, urbanisation and a growing population associated with influxes to the City as current departmental operational responsibilities (Figure 5), and a perception that responding to climate change is, very often, disconnected from day-to-day operations (Figure 6). Taylor et al.[56(p.106)] also noted similar tensions in their work on three municipalities in the country:

> *… adaptation progress to date has been reliant on coupling the climate change agenda with a dominant, pre-existing local development priority, such as market competitiveness, job creation or water security. However, when the climate change agenda conflicts with a local development priority, for example making land available for property development, it is actively marginalised, suggesting that municipal adaptation to date is limited to building resilience within traditional patterns of economic inequality and political marginalization.*

Conclusions

Cities are arguably becoming the new locus of climate change action as people move to cities for work, shelter, food and overall well-being. Cities such as Johannesburg that have emerged from past planning actions based on apartheid planning modes and are increasingly being characterised by complex informality, have highlighted that the old ways of planning may not always be fit for purpose for a city that is constantly changing. Moreover, the power differentials in cities are also notable, particularly in this case where politics, leadership and development challenges changed several times as the research progressed.

Navigating respectfully, humbly, and honestly in such a complex environment thus requires the consideration and experimentation of different approaches and methods. We chose to make use of the CHAT approach here, where the basic unit of analysis, the city climate change adaptation planning actions, provided the minimal meaningful context for understanding human action. Through the CCAF and CAP, the *object-oriented activity system*, a heterogeneous, co-engaged and multi-voiced process was enabled.[29,30] The activity system brought together the focal unit and, more critically, its social relations in a broader system (surfaced through various engagements such as Change Laboratories and smaller group interviews). In this manner, the focal unit is, and was, not seen as separate from its activities in the system, but rather was, and still is, being understood within the context and its' society/ies.[31]

Our engagements with the CoJ, as traced in Figure 4 and Table 1, were thus linked to the CHAT approach and were also strongly anchored on expansive learning processes. The four epistemic learning actions of *Questioning, Analysis, Modelling* and *Examining the new Model* were followed that have now culminated in the CAP now officially launched. The engagement of C40 through a legitimate process engaging official partners in the CoJ climate action efforts, has also given added impetus to the work we started as adaptation work in the CoJ some years ago. Several issues that we raised in our research have been taken up in the CAP. Planning action areas (over short- and long-term time scales) identified through co-designed collaborative efforts between departments have been adopted by the CAP process and further developed.[26] The beginnings of a detailed vulnerability assessment, reviewed by the research team and City officials, has also paved the way for more intensive and extensive development of vulnerability assessments that focus on social-economic, socio-cultural and biophysical hazard assessment.[26]

Several learnings, both those that are more academic and those that are more personal and reflexive, have also emerged from this work for all involved, including City officials and the academic members of the team.

The most important has been the need for a respectful, trust-building approach that is intentionally both mindful and respectful of the context in which people find themselves, both as citizens living in the City and as City officials charged with managing the City.

The 'multi-voicedness' that the CHAT approach fosters, has and continues to be especially pertinent for our work with the City – allowing us to explicitly recognise the multiple points of view and interests (both individual and collective), and the diverse histories, rules and conventions operating across the CoJ community. These voices and the eyes through which the 'system' is viewed, all set the landscape for both contradictions *and* opportunities to surface in the interactions between the different actors in the CoJ, 'demanding actions of translation and negotiation'[37(pg.136)] to facilitate shared understandings.

The implementation of the 'model solution', i.e. the CCAF and its capturing within the CAP, does not mark the end of resolving the contradictions nor probing the structural tensions that may persist when attempting to implement the City's adaption response. Rather, in fact, it marks the beginning of the emergence of *new* challenges and contradictions that arise from integrating the new solution. One of the main limitations of the study is that stakeholders such as residence associations and councillors who are affected by the City's service delivery, and the consequences of climate change impact, did not participate in the research. Their participation would have surfaced tensions and matters of concern from an 'outside' perspective, and thus, enriched both the scope and depth of the research, but also the key issues captured in the CCAF. To some extent, the absence of these voices, in particular the youth, has been remedied through the recent direct inputs into the CAP, but gaps remain. With the CAP, the City now is planning to engage on more varied actor engagements across the CoJ community.

The experiences profiled in this paper have all enabled the development of a fairly progressive CAP. The implementation of any Plan (in this case the CAP specifically) will, however, require ongoing and sustained engagement. The City and the research team remain committed to further expanding our learning so that a just and climate 'friendly' city can be handed down through the next generations.

Acknowledgements

The research team is very grateful for the time and trust gained engaging and working with the City of Johannesburg. The support and leadership actions of Lebo Molefe and her team have been, and will continue to be, extremely important as we continue to engage with the City on climate change matters of concern. We thank the anonymous reviewers for their inputs and comments that substantially improved the paper.

Competing interests

We have no competing interests to declare.

Authors' contributions

C.V.: Project leadership, conceptualisation, methodology, data collection, writing – the initial draft, funding acquisition, project management. M.M.: Conceptualisation, methodology, writing – the initial draft. K.C.: Data collection, writing – the initial draft. M.G.: Conceptualisation, funding acquisition, networking and data collection.

References

1. Carter JG, Cavan G, Connelly A, Guy S, Handley J, Kazmierczak A. Climate change and the city: Building capacity for urban adaptation. Prog Plan. 2015;95:1–66. https://doi.org/10.1016/j.progress.2013.08.001
2. Revi A, Satterthwaite D, Aragón-Durand F, Corfee-Morlot J, Kiunsi RBR, Pelling M, et al. Towards transformative adaptation in cities: The IPCC's Fifth Assessment. Environ Urban. 2014;26(1):11–28. https://doi.org/10.1177/0956247814523539
3. Roberts D, O'Donoghue S. Urban environmental challenges and climate change action in Durban, South Africa. Environ Urban. 2013;25(2):299–319. https://doi.org/10.1177/0956247813500904
4. Klein JT. Prospects for transdisciplinarity. Futures. 2004;36(4):515–526. https://doi.org/10.1016/j.futures.2003.10.007
5. Klein JT. The transdisciplinary moment (um). Integral Rev. 2013;9(2):189–99.
6. Polk M. Transdisciplinary co-production: Designing and testing a transdisciplinary research framework for societal problem solving. Futures. 2015;65:110–122. https://doi.org/10.1016/j.futures.2014.11.001
7. Engeström Y. Learning by expanding: An activity-theoretical approach to developmental research. Helsinki: Orienta-Konsultit Oy; 1987.
8. Sannino A, Engeström Y, Lemos M. Formative Interventions for expansive learning and transformative agency. J Learn Sci. 2016;25(4):599–633. https://doi.org/10.1080/10508406.2016.1204547
9. Engeström Y, Sannino A. Discursive manifestations of contradictions in organizational change efforts: A methodological framework. J Organ Change Manag. 2011;24(3):368–387. https://doi.org/10.1108/09534811111132758
10. Galderisi A. Urban resilience: A framework for empowering cities in face of heterogeneous risk factors. AZ ITU J Fac Archit. 2014;11(1):36–58.
11. The City of Johannesburg. Draft integrated development plan 2020/21 [document on the Internet]. No date [cited 2021 Feb 23]. Available from: https://www.joburg.org.za/documents_/Pages/Key%20Documents/SDBIP%20IDP%20and%20Budget/IDP%202020-21/Draft-Integrated-Development-Plan-202021.aspx
12. Bohle HG, Downing TE, Watts MJ. Climate change and social vulnerability: Toward a sociology and geography of food insecurity. Glob Environ Change. 1994;4(1):37–48. https://doi.org/10.1016/0959-3780(94)90020-5
13. Cutter SL, Mitchell JT, Scott MS. Revealing the vulnerability of people and places: A case study of Georgetown County, South Carolina. Ann Assoc Am Geogr. 2000;90(4):713–737. https://doi.org/10.1111/0004-5608.00219
14. Cutter SL, Boruff BJ, Shirley WL. Social vulnerability to environmental hazards. Soc Sci Q. 2003;84(2):242–261. https://doi.org/10.1111/1540-6237.8402002
15. Kaika M. *'Don't call me resilient again!'*: The New Urban Agenda as immunology … or … what happens when communities refuse to be vaccinated with 'smart cities' and indicators. Environ Urban. 2017;29(1):89–102. https://doi.org/10.1177/0956247816684763
16. Friend RM, Anwar NH, Dixit A, Hutanuwatr K, Jayaraman T, McGregor JA, et al. Re-imagining inclusive urban futures for transformation. Curr Opin Environ Sustain. 2016;20:67–72. https://doi.org/10.1016/j.cosust.2016.06.001
17. Mokwena L. Municipal responses to climate change in South Africa: The case of eThekwini, the City of Cape Town, and the City of Johannesburg. Report no. 113 [document on the Internet]. c2009 [cited 2020 Dec 14]. Available from: https://media.africaportal.org/documents/RR113.pdf
18. Vogel C. Review of the Climate Change Adaptation Plan for the City of Johannesburg: Unpublished report prepared for the City of Johannesburg; 2016.
19. Scott D, Davies H, New M, editors. Mainstreaming climate change in urban development: Lessons from Cape Town. Cape Town: UCT Press; 2019.
20. Chu E, Anguelovski I, Carmin J. Inclusive approaches to urban climate adaptation planning and implementation in the Global South. Clim Policy. 2016;16(3):372–392. https://doi.org/10.1080/14693062.2015.1019822
21. Ziervogel G, Parnell S. South African coastal cities: Governance responses to climate change adaptation. In: Cartwright A, Parnell S, Oelofse G, Ward S, editors. Climate change at the city scale: Impacts, mitigation and adaptation in Cape Town. London: Routledge; 2012. p. 223–243. https://doi.org/10.4324/9780203112656
22. Eisenhauer DC. Pathways to climate change adaptation: Making climate change action political: Pathways to climate change adaptation. Geogr Compass. 2016;10(5):207–221. https://doi.org/10.1111/gec3.12263
23. Davoudi S, Shaw K, Haider LJ, Quinlan AE, Peterson GD, Wilkinson C, et al. Resilience: A bridging concept or a dead end? "Reframing" resilience: Challenges for planning theory and practice interacting traps: Resilience assessment of a pasture management system in Northern Afghanistan urban resilience: What does it mean in planning practice? Resilience as a useful concept for climate change adaptation? The politics of resilience for planning: A cautionary note. Plan Theory Pract. 2012;13(2):299–333.
24. Pelling M. Adaptation to climate change: From resilience to transformation. London: Routledge; 2011.

25. Ziervogel G, Cowen A, Ziniades J. Moving from adaptive to transformative capacity: Building foundations for inclusive, thriving, and regenerative urban settlements. Sustainability. 2016;8(9):955. https://doi.org/10.3390/su8090955

26. City of Johannesburg. Climate Action Plan. Johannesburg: Environment and Infrastructure Services Department, City of Johannesburg; 2021. Available from: https://www.joburg.org.za/departments_/Documents/EISD/City%20of%20Johannesburg%20-%20Climate%20Action%20Plan%20%28CAP%29.pdf

27. Hill R, Capper P, Wilson K, Whatman R, Wong K. Workplace learning in the New Zealand apple industry network: A new co□design method for government "practice making". J Workplace Learn. 2007;19(6):359–376. https://doi.org/10.1108/13665620710777101

28. Selaolo T, Lotriet H. Towards the incorporation of activity-based learning and reflection into Botswana information systems development practice. J Workplace Learn. 2014;26(2):109–127. https://doi.org/10.1108/JWL-01-2013-0001

29. Engeström Y. Non scolae sed vitae discimus: Toward overcoming the encapsulation of school learning. Learn Instr. 1991;1(3):243–259. https://doi.org/10.1016/0959-4752(91)90006-T

30. Sawchuck PH. Adult learning and technology in working-class life. Cambridge: Cambridge University Press; 2003.

31. Ploettner J, Tresseras E. An interview with Yrjö Engeström and Annalisa Sannino on activity theory. Bellaterra J Teach Learn Lang Lit. 2016;9(4):87–98.

32. Engeström Y. Activity theory and individual and social transformation. In: Engeström Y, Miettinen R, Punamäki R-L, editors. Perspectives on activity theory. Learning in Doing: Social, Cognitive and Computational Perspectives Series. Cambridge: Cambridge University Press; 1999. p. 19–38. https://doi.org/10.1017/CBO9780511812774

33. Lotz-Sisitka H, Mukute M, Chikunda C, Baloi A, Pesanayi T. Transgressing the norm: Transformative agency in community-based learning for sustainability in southern African contexts. Int Rev Educ. 2017;63(6):897–914. https://doi.org/10.1007/s11159-017-9689-3

34. Mukute M. Developmental work research: A tool for enabling collective agricultural innovation. Wageningen: Wageningen Academic Publishers; 2015. https://doi.org/10.3920/978-90-8686-819-3

35. Engeström Y, Sannino A, Virkkunen J. On the methodological demands of formative interventions. Mind Cult Act. 2014;21(2):118–128. https://doi.org/10.1080/10749039.2014.891868

36. Virkkunen J, Newnham DS. The Change Laboratory: A tool for collaborative development of work and education. Rotterdam: Sense Publishers; 2013. https://doi.org/10.1007/978-94-6209-326-3

37. Engeström Y. Expansive learning at work: Toward an activity theoretical reconceptualization. J Educ Work. 2001;14(1):133–156. https://doi.org/10.1080/13639080020028747

38. Haapasaari A, Engeström Y, Kerosuo H. The emergence of learners' transformative agency in a Change Laboratory intervention. J Educ Work. 2016;29(2):232–262. https://doi.org/10.1080/13639080.2014.900168

39. Sannino A. The principle of double stimulation: A path to volitional action. Learn Cult Soc Interact. 2015;6:1–15. https://doi.org/10.1016/j.lcsi.2015.01.001

40. Engeström Y, Sannino A. Studies of expansive learning. In: Daniels H, editor. Introduction to Vygotsky. 3rd ed. New York: Routledge; 2017. p. 100–146. Available from: https://doi.org/10.4324/9781315647654

41. Hopwood N, Gottschalk B. Double stimulation "in the wild": Services for families with children at-risk. Learn Cult Soc Interact. 2017;13:23–37. https://doi.org/10.1016/j.lcsi.2017.01.003

42. Virkkunen J, Ristimäki P. Double stimulation in strategic concept formation: An activity-theoretical analysis of business planning in a small technology firm. Mind Cult Act. 2012;19(3):273–286. https://doi.org/10.1007/978-94-6209-326-3

43. Culwick C, Washbourne C-L, Anderson PML, Cartwright A, Patel Z, Smit W. CityLab reflections and evolutions: Nurturing knowledge and learning for urban sustainability through co-production experimentation. Curr Opin Environ Sustain. 2019;39:9–16. https://doi.org/10.1016/j.cosust.2019.05.008

44. Brugnach M, Özerol G. Knowledge co-production and transdisciplinarity: Opening Pandora's Box. Water. 2019;11(10):1997. https://doi.org/10.3390/w11101997

45. Kerosuo H, Kajamaa A, Engeström Y. Promoting innovation and learning through Change Laboratory: An example from Finnish Health care. Cent Eur J Public Policy. 2010;4:110–131.

46. Huq S. Designing adaptation projects for the Green Climate Fund. The Daily Star. 2018 May 30 [cited 2020 Dec 14]. Available from: https://www.thedailystar.net/opinion/politics-climate-change/designing-adaptation-projects-the-green-climate-fund-1583458

47. Kirchhoff CJ, Carmen Lemos M, Dessai S. Actionable knowledge for environmental decision making: Broadening the usability of climate science. Annu Rev Environ Resour. 2013;38(1):393–414. https://doi.org/10.1146/annurev-environ-022112-112828

48. Roux DJ, Nel JL, Cundill G, O'Farrell P, Fabricius C. Transdisciplinary research for systemic change: Who to learn with, what to learn about and how to learn. Sustain Sci. 2017;12(5):711–726. https://doi.org/10.1007/s11625-017-0446-0

49. Engeström Y. Innovative learning in work teams: Analyzing cycles of knowledge creation in practice. In: Punamäki R-L, Miettinen R, Engeström Y, editors. Perspectives on activity theory. Learning in Doing: Social, Cognitive and Computational Perspectives Series. Cambridge: Cambridge University Press; 1999. p. 377–404. https://doi.org/10.1017/CBO9780511812774.025

50. Mander Ü, Tournebize J, Tonderski K, Verhoeven JTA, Mitsch WJ. Planning and establishment principles for constructed wetlands and riparian buffer zones in agricultural catchments. Wetl Buffer Zones Watershed Manag. 2017;103:296–300. https://doi.org/10.1016/j.ecoleng.2016.12.006

51. Stefanakis AI. The role of constructed wetlands as green infrastructure for sustainable urban water management. Sustainability. 2019;11(24):6981. https://doi.org/10.3390/su11246981

52. Valentine G. Theorizing and researching intersectionality: A challenge for feminist geography. Prof Geogr. 2007;59(1):10–21. https://doi.org/10.1111/j.1467-9272.2007.00587.x

53. Roberts D. Thinking globally, acting locally — institutionalizing climate change at the local government level in Durban, South Africa. Environ Urban. 2008;20(2):521–537. https://doi.org/10.1177/0956247808096126

54. Roberts D. Prioritizing climate change adaptation and local level resilience in Durban, South Africa. Environ Urban. 2010;22(2):397–413. https://doi.org/10.1177/0956247810379948

55. Hetz K. Contesting adaptation synergies: Political realities in reconciling climate change adaptation with urban development in Johannesburg, South Africa. Reg Environ Change. 2016;16(4):1171–1182. https://doi.org/10.1007/s10113-015-0840-z

56. Taylor A, Cartwright A, Sutherland C. Institutional pathways for local climate adaptation: A comparison of three South African municipalities. Report no. 18. Paris: Agence Française de Développement; 2014. p. 142.

AUTHORS:
Patience T. Mulovhedzi[1]
Gift T. Rambuwani[1]
Mary-Jane Bopape[1]
Robert Maisha[1]
Nkwe Monama[2,3]

AFFILIATIONS:
[1]South African Weather Service, Pretoria, South Africa
[2]Centre for High Performance Computing (CHPC), Council for Scientific and Industrial Research, Pretoria, South Africa
[3]National Integrated Cyber-Infrastructure System (NICIS), Council for Scientific and Industrial Research, Pretoria, South Africa

CORRESPONDENCE TO:
Patience Mulovhedzi

EMAIL:
patience.mulovhedzi@weathersa.co.za

DATES:
Received: 13 July 2020
Revised: 08 Jan. 2021
Accepted: 28 Feb. 2021
Published: 29 Sep. 2021

HOW TO CITE:
Mulovhedzi PT, Rambuwani GT, Bopape M-J, Maisha R, Monama N. Model inter-comparison for short-range forecasts over the southern African domain. S Afr J Sci. 2021;117(9/10), Art. #8581. https://doi.org/10.17159/sajs.2021/8581

ARTICLE INCLUDES:
☒ Peer review
☐ Supplementary material

DATA AVAILABILITY:
☐ Open data set
☐ All data included
☒ On request from author(s)
☐ Not available
☐ Not applicable

EDITOR:
Yali Woyessa

KEYWORDS:
numerical weather prediction models, convective scale, physical process, model inter-comparison, South African Weather Service

FUNDING:
African Academy of Sciences Climate Research for Development (grant no. CR4D-19-11)

Model inter-comparison for short-range forecasts over the southern African domain

Numerical weather prediction (NWP) models have been increasing in skill and their capability to simulate weather systems and provide valuable information at convective scales has improved in recent years. Much effort has been put into developing NWP models across the globe. Representation of physical processes is one of the critical issues in NWP, and it differs from one model to another. We investigated the performance of three regional NWP models used by the South African Weather Service over southern Africa, to identify the model that produces the best deterministic forecasts for the study domain. The three models – Unified Model (UM), Consortium for Small-scale Modelling (COSMO) and Weather Research and Forecasting (WRF) – were run at a horizontal grid spacing of about 4.4 km. Model forecasts for precipitation, 2-m temperature, and wind speed were verified against different observations. Snow was evaluated against reported snow records. Both the temporal and spatial verification of the model forecasts showed that the three models are comparable, with slight variations. Temperature and wind speed forecasts were similar for the three different models. Accumulated precipitation was mostly similar, except where WRF captured small rainfall amounts from a coastal low, while it over-estimated rainfall over the ocean. The UM showed a bubble-like shape towards the tropics, while COSMO cut-off part of the rainfall band that extended from the tropics to the sub-tropics. The COSMO and WRF models simulated a larger spatial coverage of precipitation than UM and snow-report records.

Significance:

- Extreme weather events, such as tornadoes, floods, strong winds and heat waves, have significant impacts on society, the economy, infrastructure, agriculture and many other sectors. These impacts may be mitigated or even prevented through early warning systems which depend on the use of weather forecasts and information from NWP models. As South Africa depends on models from developed countries, these models may have shortcomings in capturing extreme weather events over the southern African region.

Introduction

Weather and climate impact everyday life, while extreme events can cause loss of life and injuries as well as damage to property. The impacts of adverse weather events can be reduced if effective early warnings exist. Numerical weather prediction (NWP) is an integral part of early weather warnings because it provides weather forecasters with a longer lead time than what is available in the now-casting timescale. NWP models are based on the laws of physics that govern atmospheric dynamics and thermodynamics, and they use observations as inputs to forecast the future state of the atmosphere.[1-3] This process has improved significantly over recent years due to several factors. These factors include an improved understanding and representation of physical and dynamical processes in models, advances in observation technology and data assimilation techniques, and improved computational resources and capabilities[2-4] that make it possible for NWP models to run at high resolutions like a few kilometres to hundreds of metres. The advantage of running such NWP models include improved model forecast skill such as accurate numerical prediction of near-surface weather conditions (e.g. clouds, fog, frontal precipitation) and simulation of severe weather events triggered by deep moist convection (supercell thunderstorms, intense mesoscale convective complexes, prefrontal squall line storms, tornadoes and heavy snowfall from wintertime mid-latitude cyclones)[5,6] as well as heatwaves[7]. In addition, NWP models allow for formation of structures recognisable as convective clouds and can simulate cloud microphysics such that updraught life cycle and downdraught generation are addressed fairly.[1]

Depending on several factors – such as the design of a forecasting system, model configuration, initial conditions for the model, and lateral and surface boundary conditions – different NWP models produce variable forecasts across different parts of the globe.[1,6,8] Southern Africa is characterised by numerous climatic regions that range from arid to temperate and Mediterranean winter-rainfall regions.[9-11] These regions are affected by different weather systems during different seasons, and have variable characteristics. In addition, the southern hemisphere has a higher sea area compared to land cover, which also has an effect on weather systems.[12] As a result, the simulation of these systems requires specific model parameterisation configurations which are compatible with the region, and which take into account localised influences such as land cover, soil moisture, and cloud processes.[13,14] The weather systems may vary in horizontal scale, duration, and intensity, resulting in different effects on communities. Therefore, it is critical to identify a highly skilled NWP model that can capture the weather events that affect the study domain.

Representation of physical (sub-grid) processes differs in NWP models, such that different inter-comparison studies have been developed.[8,15-17] Dabernig et al.[8] conducted a model inter-comparison study to identify the most suitable model for predicting wind power over parts of Europe. Four NWP models or configurations were used: global deterministic European Centre for Medium-Range Weather Forecasts (ECMWF-DET), regional deterministic of the Austrian weather service (ALARO), ECMWF global ensemble prediction system (ECMWF EPS), and ECMWF

global ensemble hindcast and reforecast of global ensembles (ECMWF-HC and GEFS-RF, respectively). For this comparison study, the ECMWF-DET was found to have the highest skill.

Mahlobo[17] performed a model inter-comparison of three configurations of the Unified Model (UM), namely, 12-km UM with data assimilation (DA), 12-km UM without DA and 15-km UM for a number of weather variables over South Africa. The overall results showed that the 12-km UM with DA had better and more reliable forecasts than its counterparts. Results further showed that the 15-km UM was more accurate and reliable than the 12-km without DA in simulating minimum and maximum temperatures, while the 12-km UM without DA performed better in rainfall forecasts than the 15-km UM.

The skill and capability of NWP models to provide valuable high-resolution weather information are continuously improving.[6,18] This information includes improved location and timing of weather systems.[1] High-quality weather forecasts and information are important for saving lives, protecting the environment, assisting in the prevention and mitigation of weather-related hazards, as well as preventing economic loss in agriculture, energy, and other weather-sensitive sectors.[2,19,20] Developments in convective scale NWP have been made in recent years because of improved computational resources, to the extent that a grid spacing of less than 5 km can be used for a large domain.[1,2,19] Convection-permitting models simulate deep convection explicitly, and use convection schemes either in a limited way or not at all.[1]

The United Kingdom's national weather service (Met Office) is amongst the model developers that have made massive advances in high resolution NWP modelling, with their global model running with a 10-km grid spacing. Woodhams et al.[21] conducted a convection-permitting model inter-comparison for convective storm prediction over east Africa using the UM. The study was done over a 2-year period using the 4.4-km UM convection-permitting model and the 17x25 km UM global model. The convection-permitting model performed better than the global model for sub-daily forecasts. However, within a 48-h forecast, both models showed little dependence on forecast lead time and large dependence on time of day. The authors recommended further research and consideration for ensemble forecasting.

Since 2016, the South African Weather Service (SAWS) has been using subsets from the 10-km global UM, and operationally running the UM at two convection-permitting resolutions, namely 4.4 km and 1.5 km.[22] Stein et al.[22] compared the three configurations in order to examine the benefits of increasing model resolution for forecasting convection over southern Africa. They identified benefits in using convection-permitting models in the timing of the diurnal cycle and precipitation amounts. However, the 4.4-km model showed a delayed onset of convection compared with the 1.5-km model, as well as an inconsistent bias for both convection-permitting models.

A model inter-comparison for nine NWP models, including the Weather Research and Forecasting (WRF), UM and the Consortium for Small-scale Modelling model (COSMO), was done for the simulation of the evolution of the coupled boundary layer–valley wind system.[23] The models were simulated using the same initial and boundary conditions, as well as basic physics settings. All the models depicted a similar performance for the evolution of the valley wind system, while significant differences in the simulations of the different aspects of the boundary layer and the along-valley wind were identified amongst the different models. The authors concluded that the source of differences was most likely differences in the simulated energy balance.

In this study, we investigated the performance of three NWP models – namely the UM, COSMO and WRF – over southern Africa to identify the model that produces the best forecasts for the study domain. These three models were selected for this study because they are already in use for generating operational forecasts in the Southern African Development Community (SADC) region.[12,24] The UM is run only in South Africa; however, it is the main operational model used by SAWS. COSMO and WRF are used operationally in other SADC countries including Botswana, Namibia and Tanzania. It may, however, be noted that these models are usually operationalised in the region with little to no proper evaluation of their performance.

Description of NWP models

High-resolution weather prediction models are essential for issuing suitable guidance for severe weather warnings as they can capture near-surface and small-scale severe weather events and resolve complex topography.[16,25] Each model is run at a horizontal grid spacing of 4.4 km over the SADC domain, i.e. between 5° and 55°E, and 40° and 0°S. The UM simulations were obtained from the operational simulations produced by SAWS for operational use and were initialised at 00 UTC. COSMO and WRF simulations were produced for selected case studies and were also initialised at 00 UTC start time for a 30-h simulation on the Centre for High Performance Computing server (Council for Scientific and Industrial Research, Pretoria, South Africa).[24] We were interested in analysing only the first 24 h of model forecasts, hence the choice of lead times for the COSMO and the WRF. Outputs for these models were written out hourly. The 00 UTC cycle for the UM has a 72-h forecast lead time.

The Unified Model

The UM is the name given to an atmospheric and oceanic numerical modelling software suite developed by the UK's Met Office.[26,27] The model is designed to run with both global and limited area configurations. In addition, the modelling system can also run with the atmosphere-only component and can also be coupled with a dynamic land surface and the ocean. The UM is a seamless system which can be used for prediction across various spatial scales ranging from sub-kilometre to tens of kilometres and temporal scales which range from short term to multi-decades.[27]

The radiation scheme employed was the radiative transfer by Manners et al.[28], which is suitable for use with the two-stream radiation code of Edwards and Slingo[29]. Large-scale precipitation was parameterised using a scheme based on that of Wilson and Ballard's[30] mixed-phase precipitation scheme. Light rain and drizzle falling speed were parameterised using the speed based on Abel and Shipway[31]. The warm rain processes (auto-conversion and accretion) were parameterised based on Khairoutdinov and Kogan's[32] scheme. Furthermore, the auto-conversion and accretion were bias corrected for sub-grid variability in cloud and rain water based on the parameterisations discussed in Boutle et al.[33] The ice particle distributions were parameterised using Field et al.'s[34] scheme, which calculates the microphysical transfer rates between ice and other water species.

Other parameterisations were convection, boundary layer, clouds and land surface. Convection was represented by mass flux from the turbulent alternatives convection scheme of Gregory and Rowntree[35] with the assumption of many clouds per grid box. The boundary layer was parameterised using the method described in Lock et al.[36] Clouds were parameterised using the mixed-phase scheme similarly to precipitation.[23] The separate values for cloud water and cloud ice mixing ratios were used for fractional cloud cover. The calculated cloud fraction and condensate amounts by the cloud scheme were then used as inputs to the radiation scheme.[27] Land surface was parameterised using the Joint UK Land Environment Simulator (JULES) scheme.[37,38]

The UM is run operationally as the main NWP model for short-range forecasting at SAWS. SAWS runs the UM with a horizontal grid spacing of 4.4 km and 1.5 km over southern Africa and South Africa, respectively, and these are updated four times daily, for the 00 UTC, 06 UTC, 12 UTC and 18 UTC cycles. The local configured UM obtains its initial and boundary conditions from the UK Met Office global UM model run with a horizontal grid spacing of 10 km. The 4.4-km UM, hereafter referred to as the UM, is run with 70 vertical levels over the southern African domain. The assimilation of local observations (i.e. data assimilation/DA) at SAWS had not been concluded during the time of this study, hence the models were run without DA.

The Consortium for Small-scale Modelling model

The COSMO model is a non-hydrostatic limited area atmospheric prediction model, which is formulated in a rotated geographical coordinate system.[25] The COSMO was developed by a consortium of institutions, mainly European and Asian.[25] It obtains its initial conditions and lateral boundary conditions from the Icosahedral Non-hydrostatic (ICON) global model.[5] The ICON model employs a grid spacing of 13 km globally, which allows the COSMO to be simulated at meso-β and meso-γ scales where non-hydrostatic effects are more evident in the evolution of the atmosphere.[25]

In this study, the COSMO model was run with a grid spacing of 4.4 km, 40 vertical levels and shallow convection scheme (reduced Tiedtke scheme[39] for shallow convection only). The time-step of 45 s was used for model simulations. The model runs used[39] the mass-flux convection scheme over a geographical rotated coordinate system with a generalised terrain following height coordinate system and user-defined grid stretching in the vertical[25,40]. A two-steam radiation scheme by Ritter and Geleyn[41], which accounts for short- and long-wave radiation as well as full-cloud radiation feedback, and a multi-layer soil model by Jacobesen and Heise[42], which includes snow and interception storage, were used for model runs. Numerical systems used for model runs include the Arakawa C-grid with Lorenz vertical grid staggering by Arakawa and Moorthi[43], second-order finite differences for spatial discretisation and the Runge–Kutta split explicit time integration scheme by Wicker and Skamarok[44]. Other schemes employed in the running of the model were the Flake scheme, the sea-ice scheme and the finite differencing scheme.[45-47] The orography and land cover data for running the model were obtained from the US Geological Survey.[40]

The Weather Research and Forecasting model

The WRF model serves as a back-up operational model for SAWS. The advanced research WRF model is a non-hydrostatic model.[48] The advanced research WRF model used was version 3.9.1, which was released in April 2017 (http://www2.mmm.ucar.edu/wrf/users/wrfv3.9/updates-3.9.html). The model was developed in the late 1990s through a collaborative partnership between the US National Center for Atmospheric Research (NCAR), the National Oceanic and Atmospheric Administration (represented by the US National Centers for Environmental Prediction (NCEP) and the Earth System Research Laboratory), the US Air Force, the Naval Research Laboratory, the University of Oklahoma, and the US Federal Aviation Administration.

The WRF was set up to run with a horizontal grid spacing of 4.329 km x 4.329 km using 1250x1000 grid points, with a Mercator projection applied, and is centred over the SADC region. The model topography and boundary conditions were obtained from the US Geological Survey, with a resolution of at least 2 arc minutes. The WRF is run for up to 30 h ahead, with the input data provided 3-hourly from the Global Forecast System, which is located on the University Corporation for Atmospheric Research (UCAR) Research Data Archive (RDA) webpage (https://rda.ucar.edu/datasets/ds084.1/index.html#sfol-wl-/data/) and has a horizontal grid spacing of 0.25°.

In the set-up, a time integration step was set at 15 s in order to capture high-resolution meteorological events, instead of the usual timestep (dt=6*dx). The model top was set at 40 km, with 70 vertical levels, as well as four soil levels. The physics set-up for model runs was as follows:

- WRF Single-Moment 6-class scheme, which includes ice, snow and graupel processes suitable for high-resolution simulations[49];
- New Tiedtke scheme, used previously in REGCM4 and ECMWF cy40r1 models;
- RRTMG scheme, a new version Rapid Radiative Transfer Model added from WRF version 3.1, which includes the Monte Carlo Independent Column Approximation method of random cloud overlap and major trace gases[50];
- Yonsei University scheme, Non-local-K scheme with explicit entrainment layer and parabolic K profile in unstable mixed layer[51];
- Chen-Zhang thermal roughness length over the land, which depends on vegetation height, whereby 0 is assigned for the original thermal roughness in each sfclay option;
- Noah land surface model: Unified NCEP/NCAR/AFWA scheme with soil temperature and moisture in four layers, fractional snow cover, and frozen soil physics. The modifications added improve the representation of snow and ice sheets.[52]

Validation data

A number of observational data sets were utilised to quantify the model forecasts to ensure a reliable outcome. It is essential to have high-quality and comprehensive observations to help assess and improve model performance, as well as to help communicate the level of confidence that model users should have in forecasts.[1,2]

Ground observations

SAWS operates a network of over 200 automatic weather stations from which observational data, hereafter referred to as synops, are available on an hourly or 6-hourly basis. Wind speed and surface temperature are available hourly, while accumulated rainfall and total cloud cover are available 6 hourly. Wind speed is observed at 10 m above the ground, and surface temperature is taken at 2 m above the ground. For this study, these observations are available only over South Africa, hence analysis against synops was done only over the South African domain between 15° and 35°E, and 38° and 20°S.

Satellite data

The Global Precipitation Measurement data

Global Precipitation Measurement (GPM) data are satellite-based precipitation estimates with a global coverage.[53,54] The GPM mission was launched in February 2014 as a successor for the Tropical Rainfall Measuring Mission.[55] Total precipitation estimate was downloaded in NETCDF format for this study.[53] The spacecraft used to collect GPM data has additional channels on both the dual-frequency precipitation radar and GPM microwave imager with capabilities to sense light rain and falling snow, with advanced observations of precipitation in the mid-latitudes.[54] According to Skofronick-Jackson et al.[54], GPM underestimates precipitation in the higher latitudes. The GPM data have been widely used over the African continent in order to bridge a gap in in-situ observations, which, for example, give a poor representation of observed amounts, intensities and locations of precipitation.[56-58] Suleman et al.[57] evaluated GPM data against ground observations over parts of South Africa. Their study showed that, although the data performance was variable from one location to the next, the GPM showed poor correlation with regard to rainfall magnitude and high accuracy with regard to rainfall volumes. According to Suleman et al.[57], GPM data should be used in conjunction with ground observations for more accurate results. A study to evaluate GPM data over the African continent showed that the GPM data generally agree with rain gauges, although its performance is dependent on season, region and evaluation statistics.[56] The study further showed the limitations of GPM data over areas of high topography and high performance over Lake Victoria. The GPM data set has a 30-min time interval and a spatial resolution of 0.1°.[53] For the purposes of this study, the data were merged into hourly data sets and used in conjunction with ground observations.

Snow report data

An archive of snow events observed in South Africa since the late 19th century is made available by the Snow Report community.[59] These reports indicate the geographical location for snow occurrence, as reported by the community, and not the amount of snow observed. This record assists in identifying the spatial extent to which the models were able to capture snow occurrence.

Description of case studies: High impact weather events

NWP models perform differently across different parts of the globe and for different weather phenomena. Rainfall over South Africa occurs as a result of a number of weather systems, including those that occur primarily in the tropics and mid-latitudes. This is due to South Africa's location in the subtropics. For this study, three events that occurred in 2017 were selected due to their impact on communities, infrastructure and the economy, as well as for the amount of media coverage they generated. The events occurred on 15 July 2017, 10 October 2017 and 30 December 2017, and were respectively associated with a cold front and a coastal low, a cut-off low with a ridging high, and a surface trough.

Verification

The aim of this study was to identify the strengths and weaknesses of the UM, COSMO and WRF models, in order to make decisions for producing high-quality operational forecasts and NWP data. Understanding the quality (reliability, accuracy, skill, sharpness and uncertainty) of model forecasts is useful for gaining insight into the strengths and weaknesses of the model, for model users, decision-makers and model developers.[60] For the purpose of this study, several weather variables were selected for verifying the models and evaluating their performance against different observations, namely, wind speed, surface temperature, total precipitation and accumulated snow. We highlight only a few strengths and weaknesses of the models here; more variables and further verification would be required to reach a final conclusion.

Subjective verification

The spatial distribution of snow is displayed to study how well the models' deterministic forecasts capture the spatial distribution of these weather events. These model forecasts are displayed along with the corresponding observations for eye-ball verification over the South African domain.

Objective verification

An objective verification approach was employed to measure the models' skill in predicting different weather events. Spatial distribution of model bias against ground observations and GPM measurements for surface temperature and accumulated precipitation were plotted. The bias indicates model accuracy with regard to the observations. Time series for areal average mean error (ME), root mean squared error (RMSE) and Spearman's correlation (CORR) for 2-m temperature and 10-m wind speed were calculated for each model (Equations 1–3).[61] Probability of detection (POD), false alarm rate (FAR), BIAS score and threat score (TS) were computed for 6-hourly accumulated rainfall (Equations 4–7).[61] These statistics were computed for each model at about 200 station points. These are station locations that had valid synops at the forecast hour under investigation (point-to-point verification).

$$ME = \frac{1}{n}\sum_{i=1}^{n}(f_i - o_i) \qquad \text{Equation 1}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{1}^{n}(f_i - o_i)^2} \qquad \text{Equation 2}$$

$$CORR = \frac{\sum(f - \bar{f})\sum(o - \bar{o})}{\sqrt{(f - \bar{f})^2}\sqrt{(o - \bar{o})^2}} \qquad \text{Equation 3}$$

where f is the forecast, o is the observed value, $\bar{f}$ is the mean forecast and $\bar{o}$ is the mean observed value.[61]

$$POD = \frac{hits}{hits + misses} \qquad \text{Equation 4}$$

$$FAR = \frac{false\ alarms}{hits + false\ alarms} \qquad \text{Equation 5}$$

$$BIAS = \frac{hits + false\ alarms}{hits + misses} \qquad \text{Equation 6}$$

$$TS = \frac{hits}{hits + misses + false\ alarms} \qquad \text{Equation 7}^{47}$$

Results

Cold Front:15 July 2017

The weather on 15 July 2017 was influenced by a cold front over the southwestern parts of South Africa, preceded by a coastal low along the east coast.[62] Cold temperatures with snow and heavy rainfall were observed in places over most of the western and southern parts of the country and over the Lesotho highlands.[59] A synoptic chart that indicates the observations at 12 UTC on the day can be viewed at www.weathersa.co.za/Documents/Publications/20170715.pdf. Figure 1 indicates that the 2-m temperature simulations for the three models have a similar pattern: the models generally have a cold bias. The highest cold bias for all the models is situated over the Lesotho highlands. The UM depicts the highest cold and warm bias (Figure 1a), followed by the COSMO (Figure 1b). The WRF shows the lowest cold and warm bias (Figure 1c). Most of the bias for all the models is within 5 °C of the observed temperatures.

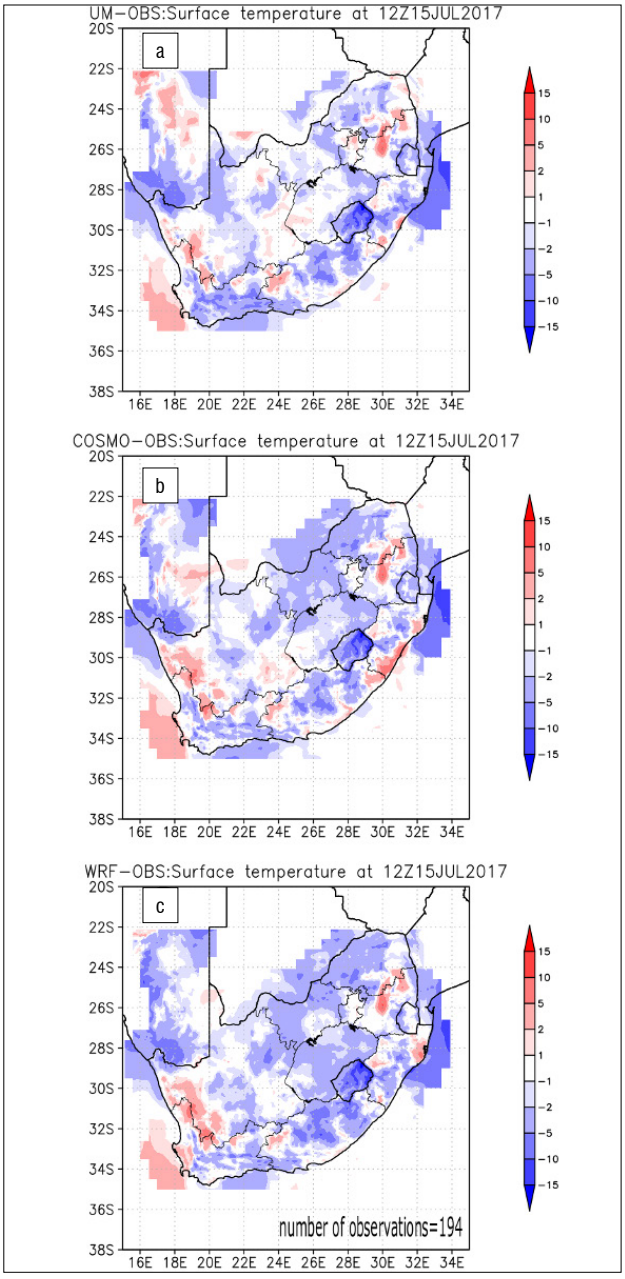


Figure 1: Surface (2-m) temperature bias over the South African domain at 1200 UTC on 15 July 2017, for the (a) UM forecast, (b) COSMO forecast and (c) WRF forecast. The bias was calculated against ground observations for 194 station points.

The precipitation that occurred on 15 July 2017 resulted from a cold front over the southwestern parts of South Africa and a coastal low ahead of it. When compared with the GPM measurements, the COSMO model depicts the highest positive bias over South Africa and adjacent oceans (Figure 2b), while the UM (Figure 2a) and WRF (Figure 2c) have a similar pattern. Figure 1a further depicts a negative bias over the southeastern parts of South Africa, implying that the UM failed to capture the precipitation that resulted from a coastal low ahead of the cold front. The COSMO and WRF depict little to no bias in the coastal low area (Figure 2b and 2c). The UM has a larger spatial coverage for positive and negative bias towards the tropics (Figure 2a). The WRF has the lowest positive and negative bias values (Figure 2c).

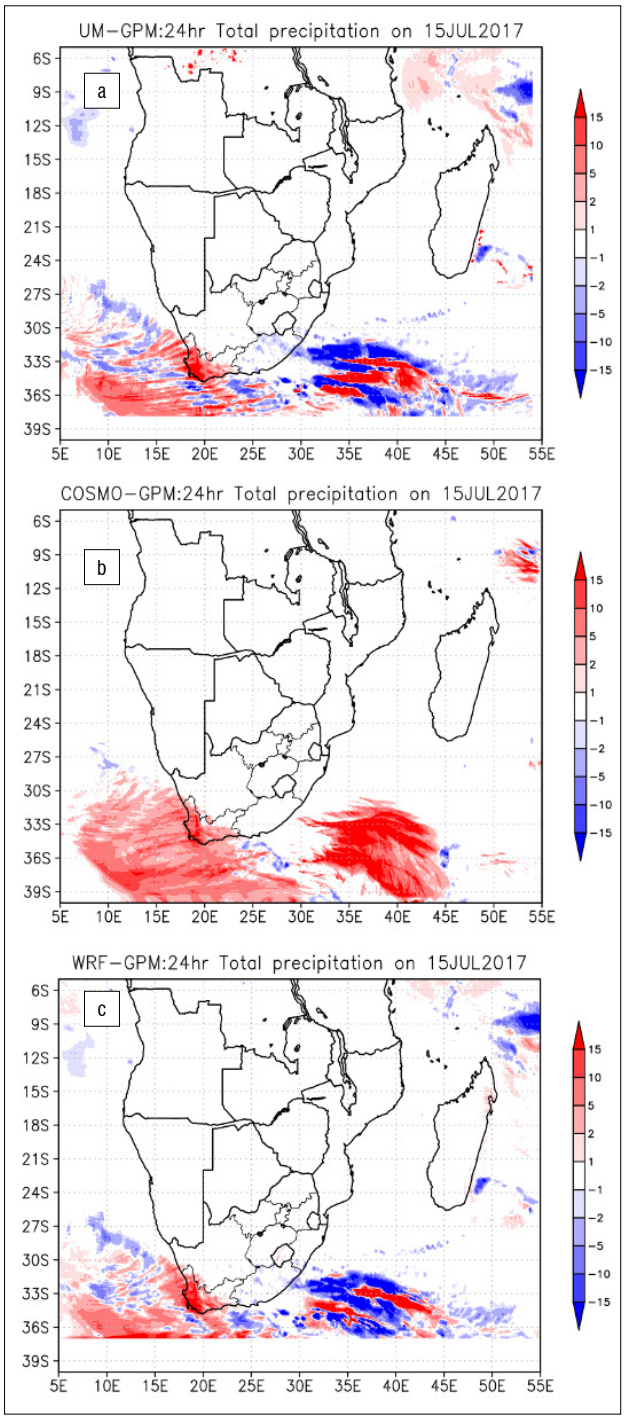


Figure 2: 24-Hour accumulated precipitation on 15 July 2017 for the South African domain where ground observations were available: (a) UM, (b) COSMO, (c) WRF, (d) ground observations and (d) GPM.

The left column in Figure 3 shows the ME, RMSE and CORR for the 2-m temperature for all three models relative to the station observations. The ME for the UM and the WRF closely correspond and generally have a small negative mean bias for most of the day, whereas the COSMO has a much higher positive bias during forecast hours 01–05, and again later in the day (Figure 3). The COSMO model shows almost no temperature bias during forecast hours 09 and 16, because the COSMO has a much larger area over South Africa where warmer temperatures over the northeastern parts of the country were over-forecast and the cooler temperatures over the southern parts were over-forecast, resulting in a small additive bias. The same applies to the early and late hours of the day when the UM and WRF have little to no bias. The UM has the lowest magnitude of errors (RMSE) in the early hours and later in the day when its RMSE is equivalent to the WRF. The RMSE for the three models is almost equivalent during the sunlight hours of the day. The WRF has the lowest correlation during the early hours, but this changes as the three models have equivalent correlation throughout the rest of the diurnal cycle. The model forecasts for 2-m temperature show high correlation with ground observations.

The right column of Figure 3 shows areal average statistics for 10-m wind speed forecasts against ground observations. The three models are comparable. The models have a negative wind speed bias throughout the diurnal cycle. The UM and COSMO have an equivalent bias (ME) throughout the day, while the WRF has lower negative bias in the early hours and higher negative bias later in the day. The models generally have large magnitudes of error (RMSE), and the UM and the COSMO correspond throughout, while the WRF is generally lower in the early hours and higher throughout the rest of the day. The models have low positive correlation for wind speed and it fluctuates throughout the day. The correlation coefficient for WRF is highest during the first two hours when the correlation coefficient for UM and COSMO are very low. The COSMO, sometimes along with the UM, shows the highest correlation coefficient during most sunlight hours, while the WRF has better performance later in the day. Wind forecast skill is poorer than temperature forecasts (Figure 3).

The skill scores for precipitation depict that the precipitation that occurred in the early hours of the day as a result of the coastal low, was only captured by the WRF (Figure 4). A cold front made landfall over the southwestern tip of South Africa later in the day, resulting in significant amounts of rainfall. Only WRF shows skill during the first half of the day, with near perfect bias scores (bias 1) at forecast hour 12 (Figure 4c). Both COSMO and WRF have higher POD during the last half of the day (Figure 4a). At the same time the three models depict equivalent FAR (Figure 4b), the COSMO and WRF show near perfect bias scores (Figure 4c) and higher TS (Figure 4d).

Ridging high and cut-off low: 10 October 2017

The weather on 10 October 2017 was characterised by a cut-off low which is an upper air disturbance that promotes uplift, coupled with a ridging high whose role is the advection of moist air from the ocean onto the land.[48] A synoptic chart that indicates the observations at 12 UTC on the day can be viewed at www.weathersa.co.za/Documents/Publications/20171010.pdf. Large parts of the southeastern half of South Africa experienced cold temperatures as a result. The models captured these temperatures; however, the extent of the area covered by lower temperatures differs across the three models. The models depict a similar pattern with areas for peak cold and warm biases located at the same points (Figure 5). UM (Figure 5a) and WRF (Figure 5c) are characterised by large areas of cold bias, while the COSMO (Figure 5b) generally over-forecasts surface temperature across the country. WRF depicts the lowest cold and warm bias compared to the other models.

Figure 6 shows the total 24-h precipitation as simulated by the three models compared to GPM rainfall estimates. The spatial coverage and intensities for the three models depict a close resemblance when compared with the GPM measurements (Figure 6). The UM shows the highest positive and negative bias (Figure 6a), followed by COSMO (Figure 6b). The WRF generally depicts the lowest positive and negative bias (Figure 6c).

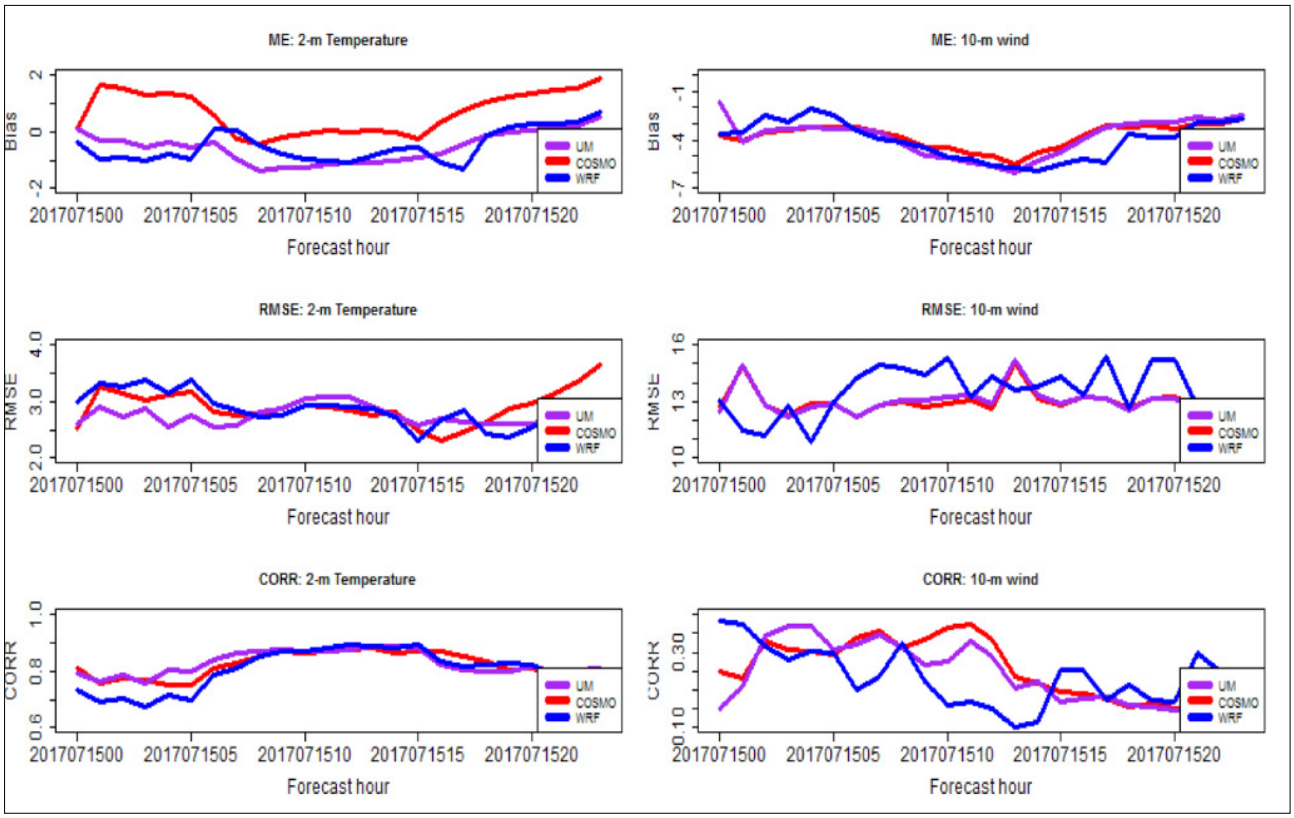


Figure 3: Diurnal cycle for areal average temperature evaluation statistics (left column) and for areal average wind speed evaluation statistics (right column) on 15 July 2017. The models – UM (purple), COSMO (red) and WRF (blue) – were evaluated against ground observations at station points that had valid data for the hour of interest.

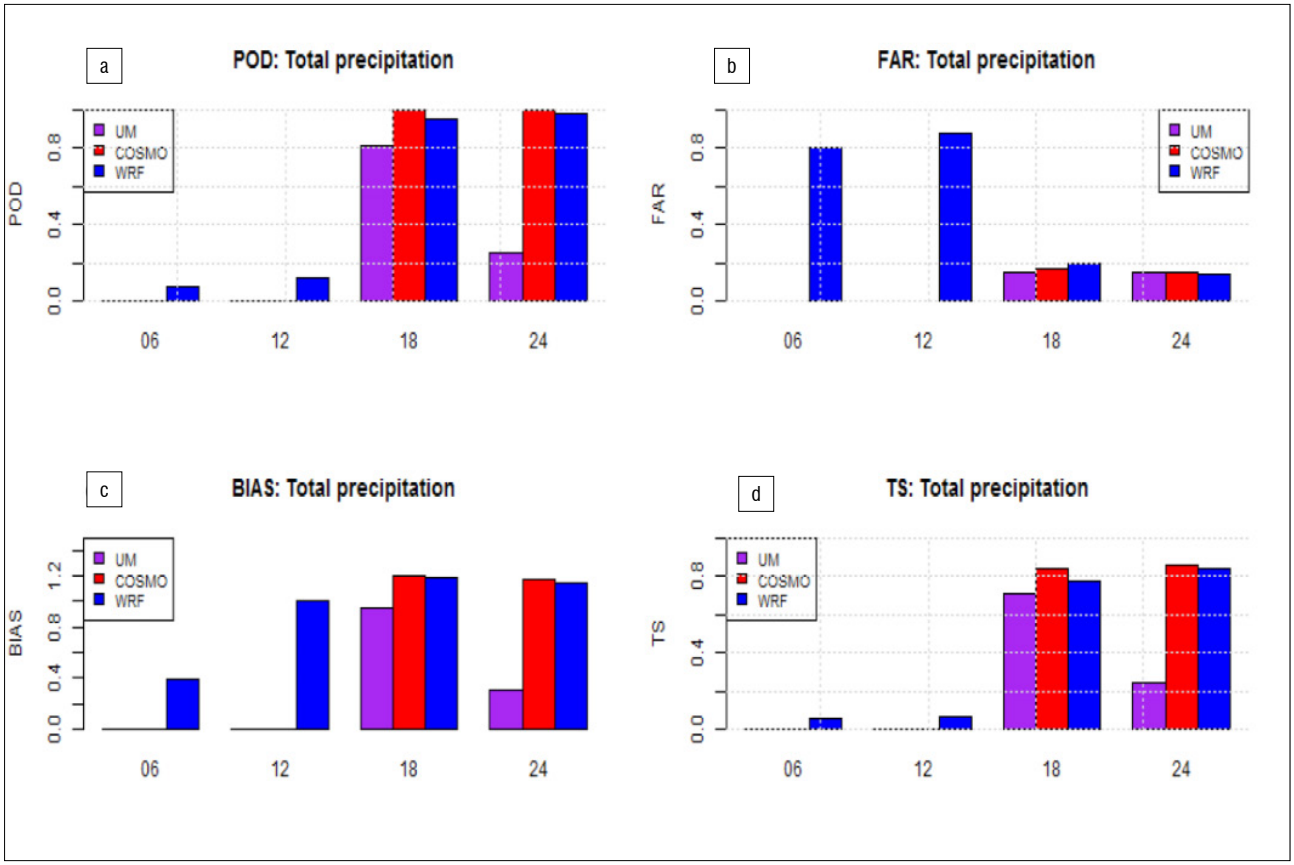


Figure 4: (a) Probability of detection (POD), (b) false alarm rate (FAR), (c) bias and (d) threat score (TS) for 6-hourly accumulated precipitation on 15 July 2017. The models: UM (purple), COSMO (red) and WRF (blue), were evaluated against ground observations at station points that had valid data for the hour of interest.

All the models captured the northwest to southeast rainfall pattern, while underestimating the amount in some areas. The rainfall over the southeastern part of South Africa was captured by all the models; however, WRF (Figure 6c) extended the rainfall area more south and north along the coastal area. It may be noted that this event resulted in flooding over the east coast of South Africa, caused severe damage to property and eight people lost their lives.

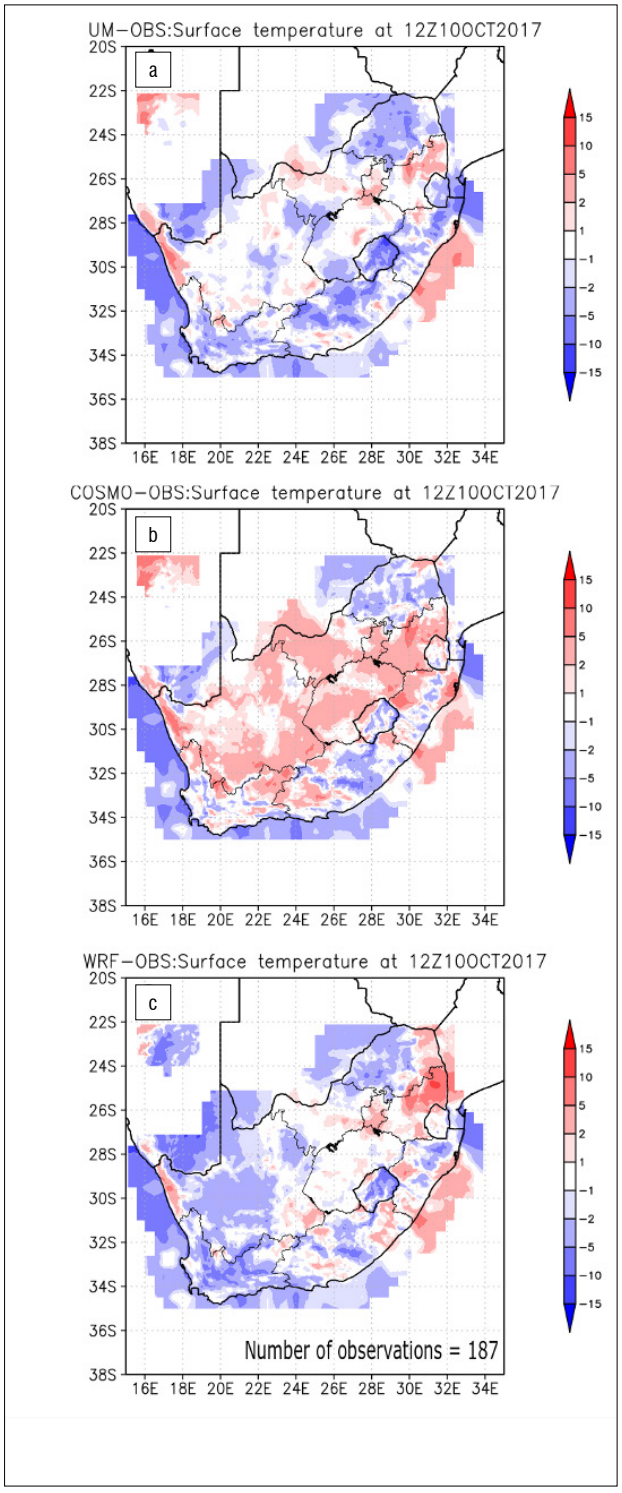


Figure 5: Surface (2-m) temperature bias over the South African domain at 1200 UTC on 10 October 2017: (a) UM forecast, (b) COSMO forecast and (c) WRF forecast. The bias was calculated against ground observations at 187 station points.

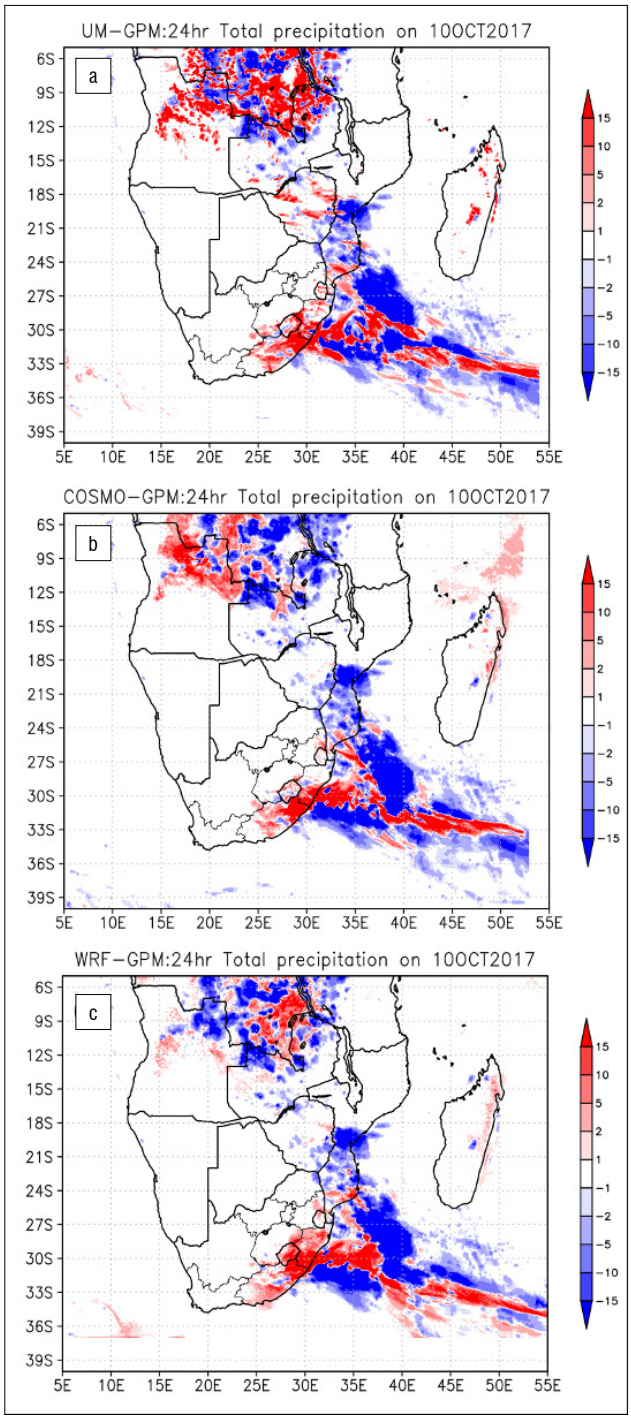


Figure 6: Accumulated precipitation amounts for a 24-h period on 10 October 2017: (a) UM forecast, (b) COSMO forecast, (c) WRF forecast, (d) GPM measurements and (e) Meteorologix satellite measurements.

Surface trough: 30 December 2017

The weather on 30 December 2017 was characterised by a broad surface trough that extended from the central interior to the western parts of South Africa, with a high to the east of the country.[60] This surface trough resulted in severe thunderstorms associated with a tornado over the central interior of South Africa. A synoptic chart that indicates the observations at 12 UTC on the day can be viewed at www.weathersa.co.za/Documents/Publications/20171230.pdf. The observations at 00 UCT on 30 December 2017 were characterised by cooler temperatures

across most of the eastern and southern parts of the country. A warm tongue was observed in the Northern Cape which extends into Namibia. The three models were able to capture the general spatial pattern showing lower temperatures over the south and east of the country. The models' simulations depict a similar pattern for surface temperature simulations with peaks for cold and warm biases situated at the same locations (Figure 7). The UM and COSMO generally show a positive bias (Figure 7a,b). WRF depicts larger areas of cold bias within 5 °C.

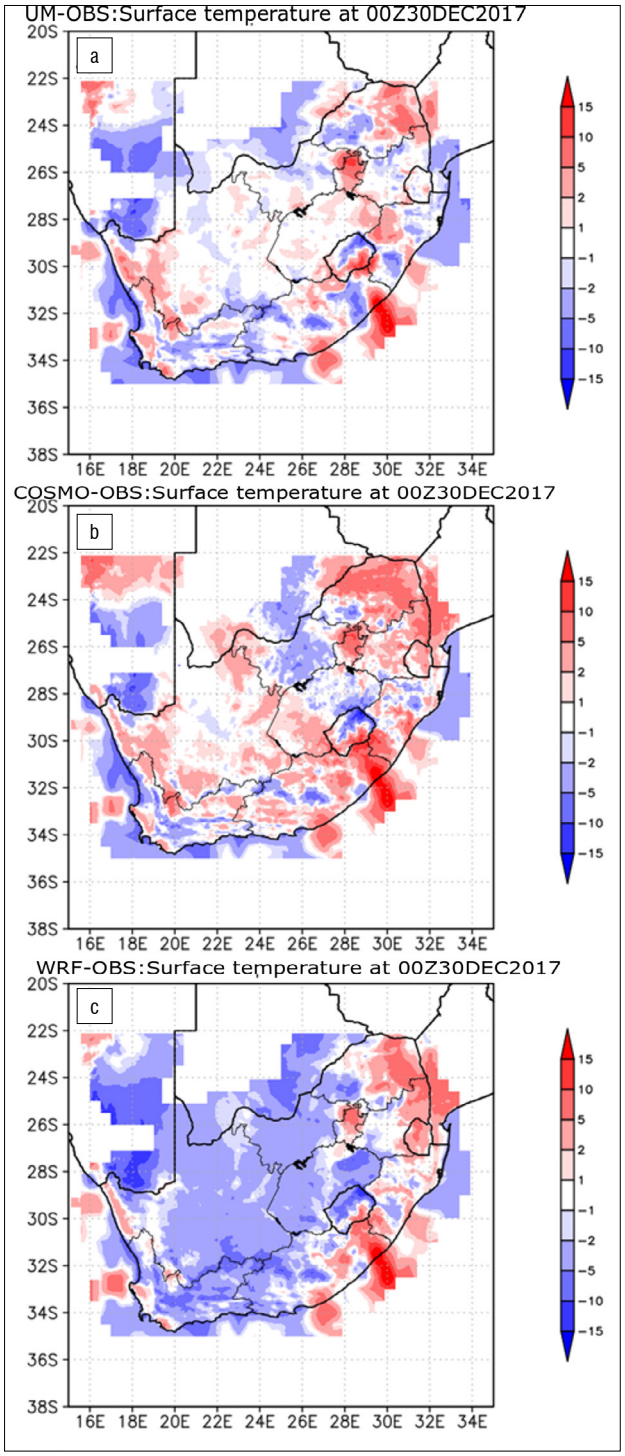


Figure 7: Surface (2-m) temperature bias over the South African domain at 0000 UTC on 30 December 2017: (a) UM forecast, (b) COSMO forecast and (c) WRF forecast. The bias was calculated against ground observations at 171 station points.

A large band of rainfall, extending from the tropics to the southern coast of South Africa, was observed on 30 December 2017. The 6-hourly accumulated precipitation for models compared with ground observations were used in this case as they contain more detail than the 24-hour totals. All the time steps followed a similar pattern, hence one time step is depicted. The models depict a similarity in simulating precipitation (Figure 8). The UM depicts the highest positive and negative biases, followed by WRF.

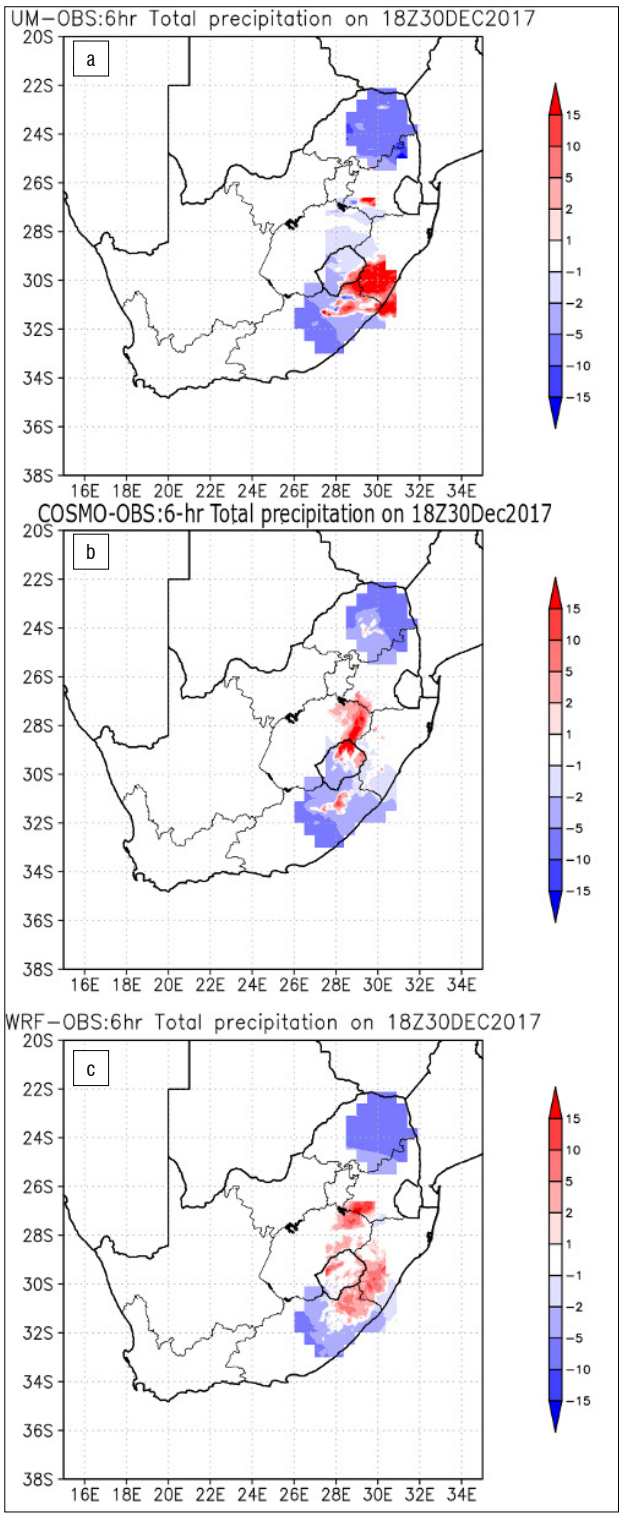


Figure 8: Bias for accumulated precipitation amounts for a 6-h period on 30 December 2017, 18 GMT: (a) UM forecast, (b) COSMO forecast and (c) WRF forecast. The bias was calculated against observations at 185 station points.

Additive bias for 2-m temperature shows high accuracy for the UM throughout the day (Figure 9, left column). There is a general over-forecast by the COSMO throughout the day. The WRF shows an over-forecast in the early and late hours, and a gradual decline during sunlight hours, to as low as -1 towards midday. The models have equivalent magnitudes of error (RMSE) for temperature throughout the diurnal cycle. The correlation coefficient for the models is also equivalent, although it is low.

The right column of Figure 9 shows statistics for 10-m wind speed. The models under-forecast wind speed throughout the diurnal cycle, and they have equivalent ME, except during the warm hours of the day when the COSMO simulation is relatively more accurate. The models show equivalent RMSE throughout the diurnal cycle. The models show a low correlation throughout the diurnal cycle.

The POD for precipitation at station points over South Africa is higher for the COSMO during the first half of the day, and higher for the WRF for the rest of the day (Figure 10). The WRF has the highest FAR at forecast hours 06 and 18, while the COSMO has the highest at forecast hour 12 and UM at forecast hour 24. The UM under-forecasts precipitation throughout the diurnal cycle. The WRF shows relatively high accuracy throughout (bias closest to the perfect score of 1); although the models have low TS throughout the diurnal cycle, WRF shows better skill compared to the other models.

Discussions and conclusions

In this study, we evaluated the performance of three models, used within the SADC region for operational weather forecasting in South Africa, in simulating three high-impact weather events. These three events were associated with a cold front, a ridging high associated with an upper-air cut-off low, and a surface trough and tornado, respectively. The second event resulted in flooding over parts of the east coast of South Africa with eight people reported to have lost their lives. The third event was reported to have damaged houses in parts of the central interior of South Africa. All three models were able to capture the events, with slight differences in performance.

All the models were able to capture the general temperature pattern with an acceptable areal average bias of about 1 °C. COSMO generally underestimated the size of the area associated with lower temperatures. The three models generally underestimated the temperature in the Limpopo Province, especially in the western parts of the province. The models failed to capture warmer temperatures over the western interior of South Africa that extended into Namibia. WRF captured this feature best; however, the area was overestimated. The three models poorly simulated 10-m winds, and as a result produced high bias and poorly correlated with ground observations.

The three models were able to capture the general rainfall pattern associated with the three case studies: 24-h accumulated precipitation for the three models was comparable to the observations, while 6-h accumulated rainfall showed poor performance. This is most noticeable on 30 December 2017 where FAR is high and POD is poor. The spatial pattern of the rainfall differed across the three models and, in some instances, the WRF and the COSMO underestimated the rainfall amount. The COSMO did not capture all the rainfall associated with the rain bands that extended from the tropics to South Africa in a northwest to southeast pattern on 10 October 2017 and 30 December 2017.

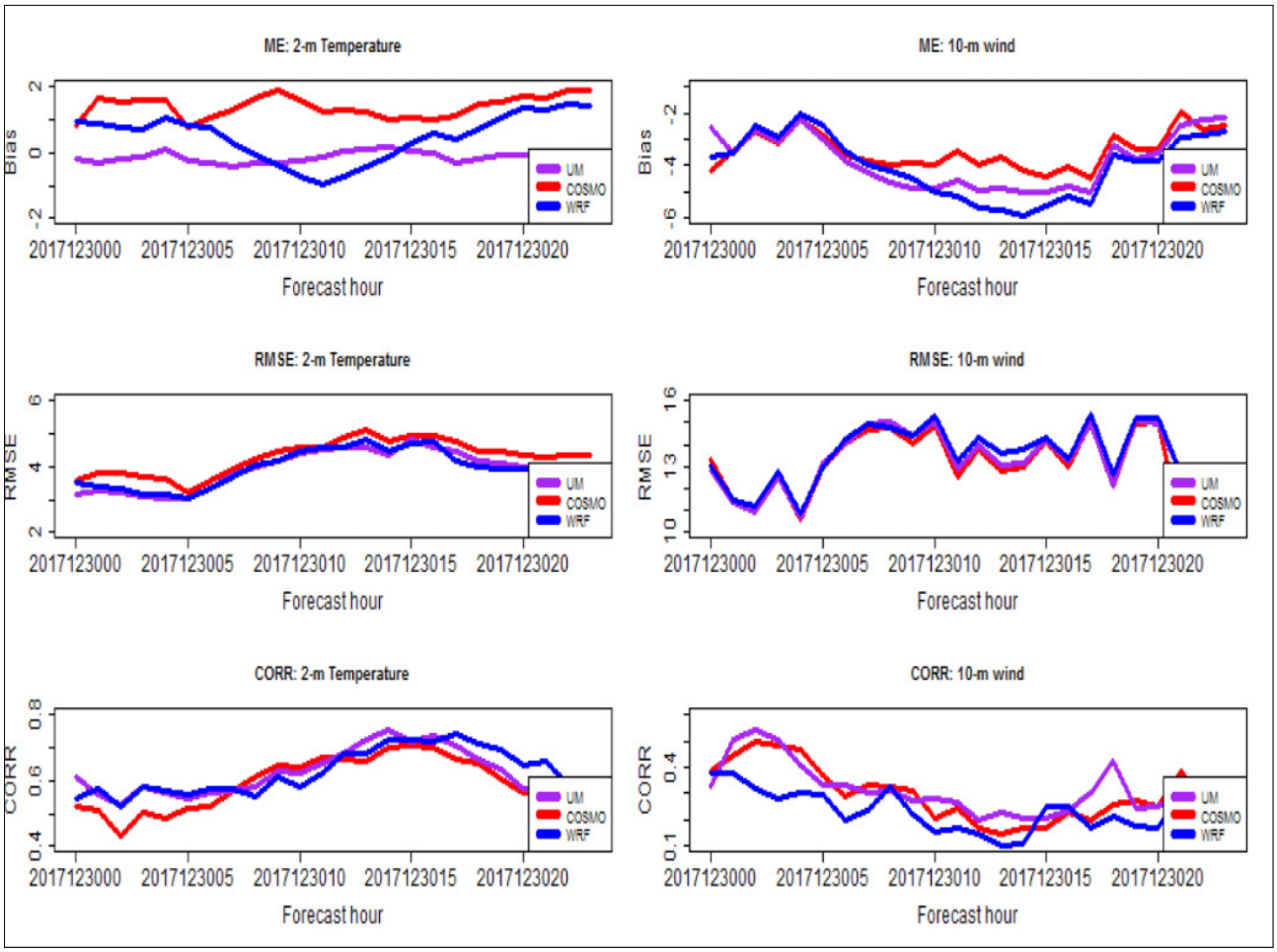


Figure 9: Diurnal cycle for areal average temperature evaluation statistics (left column) and for areal average wind speed evaluation statistics (right column) on 30 December 2017. The models – UM (purple), COSMO (red) and WRF (blue) – were evaluated against ground observations at station points that had valid data for the hour of interest.

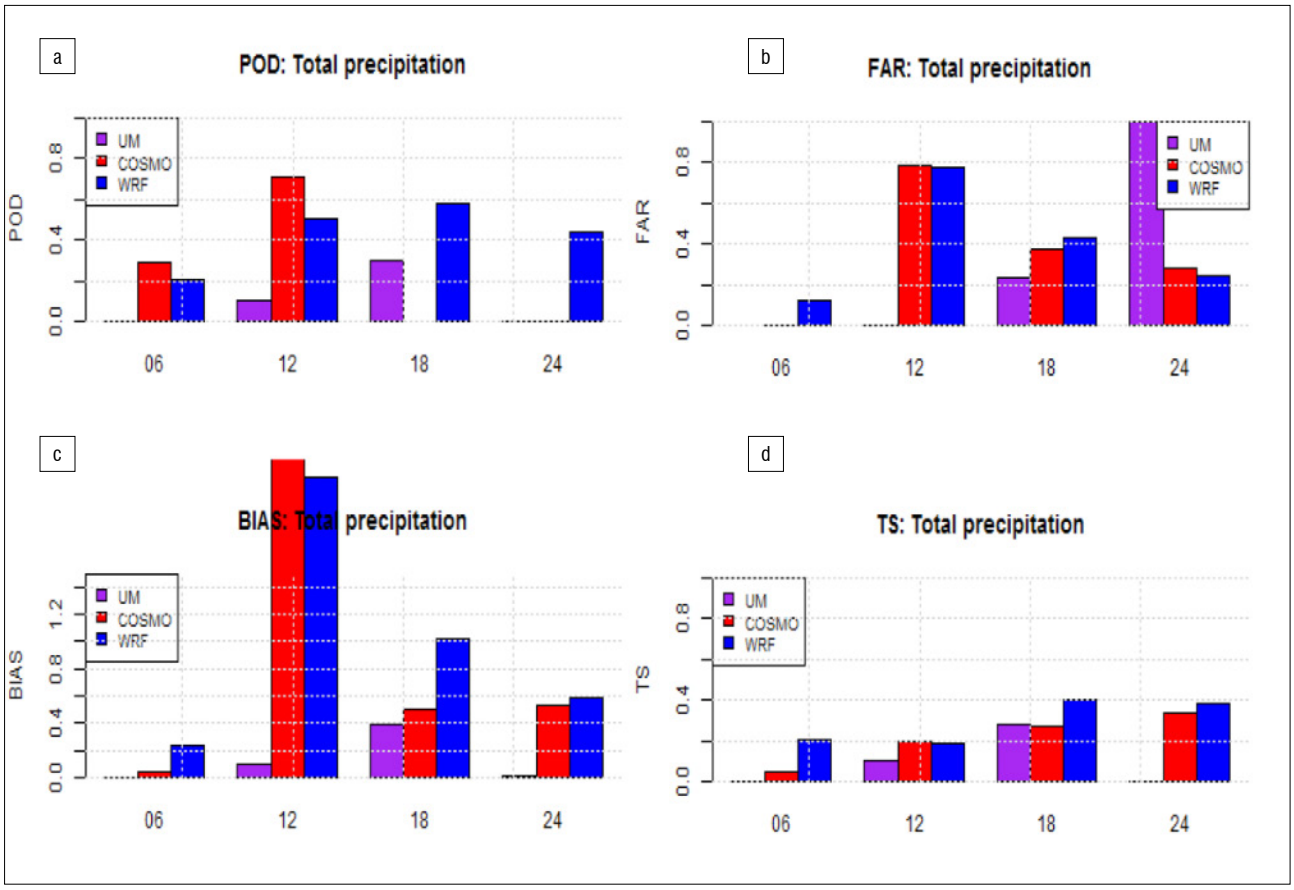


Figure 10: (a) Probability of detection (POD), (b) false alarm rate (FAR), (c) bias and (d) threat score (TS) for 6-hourly accumulated precipitation on 15 July 2017. The models – UM (purple), COSMO (red) and WRF (blue) – were evaluated against ground observations at station points that had valid data for the hour of interest.

WRF was able to capture rainfall associated with the coastal low better than the other two models. COSMO captured the event, but underestimated the rainfall, while the UM did not capture the event at all. The UM generally depicted a similar or higher amount of rainfall as that observed, but had a much smaller spatial coverage (blobbiness). COSMO generally failed to capture small and residual rainfall, and also forecasted slightly lower amounts of rainfall than the GPM observations. WRF performed well over land, and even captured the small (and residual) rainfall that the UM and COSMO did not capture. However, it overestimated rainfall over the ocean.

The study shows that the models are skilled in capturing general details of big events, similar to the ones that were analysed here. The differences in the simulations point to issues with small-scale processes and, in particular, in the simulation of rainfall. The poor performance of the models may be associated with the use of coarse-resolution observations and non-gridded data sets.[1,63] It is important that further research is conducted to understand the reasons associated with the different model performances. However, it may be noted, as also shown in this study, that there are shortcomings in the available observations, which makes it difficult for models to be verified in detail. Satellite estimates show different amounts of rainfall, as also shown here.

Acknowledgements

We are grateful for a Climate Research for Development (CR4D) fellowship grant (reference number CR4D-19-11), managed by the African Academy of Sciences. We also thank the South African National Integrated Cyber-Infrastructure System (NICIS), Centre for High Performance Computing (CHPC), for the computational resources as well as model set-up, and Germany's national meteorological service, the Deutscher Wetterdienst, for helping with the COSMO model set-up.

Competing interests

We have no competing interests to declare.

Authors' contributions

P.T.M.: Conceptualisation, methodology, data collection, data analysis and validation, data curation, writing – the initial draft. G.T.R.: Methodology, data analysis and validation, data curation, writing – the initial draft. M-J.B.: Methodology, data collection, data analysis and validation, data curation, funding acquisition, writing – the initial draft. R.M.: Writing – the initial draft. N.M.: Data curation.

References

1. Clark P, Roberts N, Lean H, Ballard SP, Charlton-Perez C. Convection-permitting models: A step-change in rainfall forecasting. Meteorol Appl. 2016;23:165–181. https://doi.org/10.1002/met.1538
2. Bauer P, Thorpe A, Brunet G. The quiet revolution of numerical weather prediction. Nature. 2015;525:47–55. https://doi.org/10.1038/nature14956
3. Schultze GC. Atmospheric observations and numerical weather prediction. S Afr J Sci. 2007;103(7–8):318–323.
4. Duan Q, Di G, Quan J, Wang C, Gong W. Automatic model calibration: A new way to improve numerical weather forecasting. Bull Am Meteorol Soc. 2017;98(5):959–970. https://doi.org/10.1175/BAMS-D-15-00104.1
5. Schättler U, Doms U, Schraff C. A Description of the nonhydrostatic regional COSMO-model. Offenbach: Deutscher Wetterdienst; 2014.
6. Tang Y, Lean WH, Bornemann J. The benefits of the Met Office variable resolution NWP model for forecasting convection. Meteorol Appl. 2013;20:417–426. https://doi.org/10.1002/met.1300

7. Maisha TR. The influence of topography and grid resolution on extreme weather forecasts over South Africa [thesis]. Pretoria: University of Pretoria; 2014.

8. Dabernig M, Mayr GJ, Messner JW. Predicting wind power with reforecasts. Weather Forecast. 2015;30:1655–1662. https://doi.org/10.1175/WAF-D-15-0095.1

9. Peel MC, Finlayson BL, McMahon TA. Updated world map of the Köppen-Geiger climate classification. Hydrol Earth Syst Sci. 2007;11:1633–1644. https://doi.org/10.5194/hess-11-1633-2007

10. Daron JD. Regional climate messages: Southern Africa. Scientific report from the CARIIA Adaptation at Scale in Semi-Arid Regions (ASSAR) Project. Ottawa: CARIAA and ASSAR; 2014.

11. Engelbrecht CJ, Engelbrecht FA. Shifts in Köppen-Geiger climate zones over southern Africa in relation to key global temperature goals. Theor Appl Climatol. 2016;123:247–261. https://doi.org/10.1007/s00704-014-1354-1

12. Dong B, Gregory JM, Sutton RT. Understanding land-sea warming contrast in response to increasing greenhouse gases. Part I: Transient adjustment. J Clim. 2009;22:3079–3097. https://doi.org/10.1175/2009JCLI2652.1

13. Liou KN. Radiation and cloud processes in the atmosphere: Theory, observation and modeling. London: Oxford University Press; 1992.

14. Sellers PJ. Biophysical models of land surface processes. In: Trenberth K, editor. Climate system modeling. Cambridge, UK: Cambridge University Press; 1993.

15. Krishnamurti TN, Bounoua L. An introduction to numerical weather prediction techniques. Boca Raton, FL: CRC Press; 1996. https://doi.org/10.1201/9781315137285

16. Šaur D. Evaluation of the accuracy of numerical weather prediction models. In: Silhavy R, Senkerik R, Oplatkova Z, Prokopova Z, Silhavy P, editors. Artificial intelligence perspectives and applications. Advances in intelligent systems and computing vol 347. Cham: Springer; 2015. https://doi.org/10.1007/978-3-319-18476-0_19

17. Mahlobo DD. The verification of different model configurations of the Unified Atmospheric Model over South Africa [thesis]. Pretoria: University of Pretoria; 2013.

18. Kann A, Wittmann C, Bica B, Wastl C. On the impact of NWP model background on very high-resolution analysis in complex terrain. Weather Forecast. 2015;30(4):1077–1089. https://doi.org/10.1175/WAF-D-15-0001.1

19. Baldauf M, Seifert A, Forster J, Majewski D, Raschendorfer M, Reinhardt T. Operational convective-scale numerical weather prediction with the COSMO model: Description and sensitivities. Mon Weather Rev. 2011;139(12):3887–3904. https://doi.org/10.1175/MWR-D-10-05013.1

20. Mulovhedzi P. Towards a heat-watch warning system for South Africa for the benefit of the health sector [thesis]. Pretoria: University of Pretoria; 2017.

21. Woodhams BB. What is the added value of a convection-permitting model for forecasting extreme rainfall over tropical East Africa? Mon Weather Rev. 2018;146:2757–2780. https://doi.org/10.1175/MWR-D-17-0396.1

22. Stein THM, Keat W, Maidment RI, Landman S, Becker E, Boyd DFA, et al. An evaluation of clouds and precipitation in convection-permitting forecasts for South Africa. Weather Forecast. 2019;34(1):233–254. https://doi.org/10.1175/WAF-D-18-0080.1

23. Schmidli J, Billings B, Chow F, De Wekker S, Doyle J, Grubišić V, et al. Intercomparison of mesoscale model simulations of the daytime valley wind system. Mon Weather Rev. 2011;139:1389–1409. https://doi.org/10.1175/2010MWR3523.1

24. Bopape M-JM, Sithole HM, Motshegwa T, Rakate E, Engelbrecht F, Archer E, et al. A regional project in support of the SADC Cyber-Infrastructure Framework Implementation: Weather and climate. Data Sci J. 2019;18:1–10. https://doi.org/10.5334/dsj-2019-034

25. Doms G, Baldauf M. A description of the nonhydrostatic regional COSMO-model. Part I: Dynamics and numerics. Offenbach: Deutscher Wetterdienst; 2018.

26. Davies T, Cullen MJP, Malcolm AJ, Mawson MH, Staniford A, White AA, et al. A new dynamical core for the Met Office's global and regional modelling of the atmosphere. Q J Roy Meteor Soc. 2005;131:1759–1782. https://doi.org/10.1256/qj.04.101

27. Greed G, Haywood JM, Milton S, Keil A, Christopher S, Gupta P, et al. Aerosol optical depths over North Africa 2: Modeling and model validation. J Geophys Res. 2008;113:D00C05. https://doi.org/10.1029/2007JD009457

28. Manners J, Vosper SB, Roberts N. Radiative transfer over resolved topographic features for high-resolution weather prediction. Q J Roy Meteor Soc. 2012;138:720–733. https://doi.org/10.1002/qj.956

29. Edwards JM, Slingo A. Studies with a flexible new radiation code. I: Choosing a configuration for a largescale model. Q J Roy Meteor Soc. 1996;122:689–719. https://doi.org/10.1002/qj.49712253107

30. Wilson DR, Ballard SP. A micro-physically based precipitation scheme for the UK Meteorological Office Unified Model. Q J Roy Meteor Soc. 1999;125:1607–1636. https://doi.org/10.1002/qj.49712555707

31. Abel SJ, Shipway BJ. A comparison of cloud-resolving model simulations of trade wind cumulus with aircraft observations taken during RICO. Q J Roy Meteor Soc. 2007;133:781–794. https://doi.org/10.1002/qj.55

32. Khairoutdinov M, Kogan Y. A new cloud physics parameterization in a large-eddy simulation model of marine stratocumulus. Mon Weather Rev. 2000;128:229–243. https://doi.org/10.1175/1520-0493(2000)128<0229:ANCPPI>2.0.CO;2

33. Boutle IA, Abel SJ, Hill PG, Morcrette CJ. Spatial variability of liquid cloud and rain: Observations and microphysical effects. Q J Roy Meteor Soc. 2014;140:583–594. https://doi.org/10.1002/qj.2140

34. Field PR, Hogan RJ, Brown PRA, Illingworth AJ, Choularton TW, Cotton RJ. Parametrization of ice-particle size distributions for mid-latitude stratiform cloud. Q J Roy Meteor Soc. 2005;13:1997–2017. https://doi.org/10.1256/qj.04.134

35. Gregory D, Rowntree PR. A mass-flux convection scheme with representation of cloud ensemble characteristics and stability dependent closure. Mon Weather Rev. 1990;118:1483–1506. https://doi.org/10.1175/1520-0493(1990)118<1483:AMFCSW>2.0.CO;2

36. Lock AP, Brown AR, Bush MR, Martin GM, Smith RNB. A new boundary layer mixing scheme. Part I: Scheme description and single-column model tests. Mon Weather Rev. 2000;128:3187–3199. https://doi.org/10.1175/1520-0493(2000)128<3187:ANBLMS>2.0.CO;2

37. Best MJ, Pryor M, Clark DB, Rooney GG, Essery RLH, Ménard CB, et al. The Joint UK Land Environment Simulator (JULES), model description – Part 1: Energy and water fluxes. Geosci Model Dev. 2011;4:677–699. https://doi.org/10.5194/gmd-4-677-2011

38. Clark DB, Mercado LM, Sitch S, Jones CD, Gedney N, Best MJ, et al. The Joint UK Land Environment Simulator (JULES), model description – Part 2: Carbon fluxes and vegetation dynamics. Geosci Model Dev. 2011;4:701–722. https://doi.org/10.5194/gmd-4-701-2011

39. Tiedtke M. A comprehensive mass flux scheme for cumulus parameterization in large-scale models. Mon Weather Rev. 1989;117:1779–1799. https://doi.org/10.1175/1520-0493(1989)117<1779:ACMFSF>2.0.CO;2

40. Doms G, Förstner J, Heise E, Herzog H-J, Mironov D, Raschendorfer M, et al. A description of the nonhydrostatic regional COSMO model. Part II: Physical parameterization. Offenbach: Deutscher Wetterdienst; 2018.

41. Ritter B, Geleyn J-F. A comprehensive radiation scheme for numerical weather prediction models with potential applications in climate simulations. Mon Weather Rev. 1992;120:303–325. https://doi.org/10.1175/1520-0493(1992)120<0303:ACRSFN>2.0.CO;2

42. Jacobsen I, Heise E. A new economic method for the computation of the surface temperature in numerical models. Contr Atmos Phys.1982;55:128–141.

43. Arakawa A, Moorthi S. Baroclinic instability in vertically discrete systems. J Atmos Sci. 1988;45(11):1688–1708. https://doi.org/10.1175/1520-0469(1988)045<1688:BIIVDS>2.0.CO;2

44. Wicker L, Skamarock W. Time-splitting methods for elastic models using forward time schemes. Mon Weather Rev. 2002;130:2088–2097. https://doi.org/10.1175/1520-0493(2002)130<2088:TSMFEM>2.0.CO;2

45. Mironov D. Parameterization of lakes in numerical weather prediction. Description of a lake model. COSMO Technical report no. 11. Offenbach: Deutscher Wetterdienst; 2008.

46. Mironov D, Ritter B. Testing the new ice model for the global NWP system GME of the German Weather Service. In: Cote J, editor. Research activities in atmospheric and oceanic modelling. Rep. 34 WMO/TD 1220 4.21-4.22. Geneva: World Meteorological Organization; 2004.

47. Schulz J-P. Introducing a sea ice scheme in the COSMO model. COSMO Newsletter. 2011;11:32–40.

48. Skamarock WC, Klemp JB, Dudhia J, Gill DO, Barker DM, Duda MG, et al. A description of the advanced research WRF version 3. NCAR Technical note NCAR/TN-475+STR. Boulder, CO: Mesoscale and Microscale Meteorology Division, National Center for Atmospheric Research; 2008. https://doi.org/10.5065/D68S4MVH

49. Hong SY, Lim OJ. The WRF single-moment 6-class microphysics scheme (WSM6). J Korean Meteor Soc. 2006;42:129–151.

50. Iacono MJ, Delamere JS, Mlawer EJ, Shephard MW, Clough SA, Collins WD. Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. J Geophys Res. 2008;113:D13103. https://doi.org/10.1029/2008JD009944

51. Hong S-Y, Noh Y, Dudhia J. A new vertical diffusion package with an explicit treatment of entrainment processes. Mon Weather Rev. 2006;134:2318–2341. https://doi.org/10.1175/MWR3199.1

52. Skamarock WC, Klemp JB, Dudhia J, Gill DO, Liu Z, Berner J, et al. A Description of the advanced research WRF version 4. NCAR Technical note NCAR/TN-556+STR. Boulder, CO: Mesoscale and Microscale Meteorology Division, National Center for Atmospheric Research; 2019.

53. US National Aeronautics and Space Administration (NASA). Global Precipitation Measurement, GPM Data Downloads [webpage on the Internet]. No date [cited 2018 Jul 20]. Available from: https://gpm.nasa.gov/data-access/downloads/gpm

54. Skofronick-Jackson G, Petersen WA, Berg W, Kidd C, Stocker EF, Kirschbaum DB, et al. The Global Precipitation Measurement (GPM) Mission for science and society. Bull Am Meteorol Soc. 2017; 98:1679–1695. https://doi.org/10.1175/BAMS-D-15-00306.1

55. Retalis A, Katsanos D, Tymvios D, Michaeli S. Comparison of GPM IMERG and TRMM 3B43 products over Cyprus. Remote Sens. 2020;12:3212. https://doi.org/10.3390/rs12193212

56. Dezfuli AK, Ichoku CM, Huffman GJ, Mohr KI, Selker JS, Van de Giesen N, et al. Validation of IMERG precipitation in Africa. J Hydrometeorol. 2017;18:2817–2825. https://doi.org/10.1175/JHM-D-17-0139.1

57. Suleman S, Chetty KT, Clark DJ, Kapangaziwiri E. Assessment of satellite-derived rainfall and its use in the ACRU agro-hydrological model. Water SA. 2020;46(4):547–557. https://doi.org/10.17159/wsa/2020.v46.i4.9068

58. Muofhe T, Chikoore H, Bopape M-J, Nethengwe N, Ndarana T, Rambuwani GT. Forecasting intense cut-off lows in South Africa using the 4.4 km Unified Model. Climate. 2020;8:129. https://doi.org/10.3390/cli8110129

59. Snow Report SA. The history of snow in southern Africa [webpage on the Internet]. No date [cited 2019 June 24]. Available from: https://snowreport.co.za/history-of-snow-in-south-africa-after-2010

60. Gordon N, Shaykewich J. Guidelines on performance assessment of public weather services. WMO/TD No.1023. Geneva: World Meteorological Organization; 2000.

61. Sass BH, Yang X. A verification score for high resolution NWP: Idealized and preoperational tests. HIRLAM Technical report no. 69. De Bilt: KNMI; 2012.

62. South African Weather Service (SAWS). Caelum: Notable weather and weather-related events in 2017. Pretoria: SAWS; 2018.

63. Champion AJ, Hodges K. Importance of resolution and model configuration when downscaling extreme precipitation. Tellus A. 2014;66:23993. https://doi.org/10.3402/tellusa.v66.23993

AUTHORS:
Reena H. Seebocus[1]
Michel R. Lollchund[1]
Miloud Bessafi[2]

AFFILIATIONS:
[1]Department of Physics, University of Mauritius, Réduit, Mauritius
[2]Energy, Electronic and Processes Laboratory, University of La Réunion, Saint-Denis, Réunion

CORRESPONDENCE TO:
Michel Lollchund

EMAIL:
r.lollchund@uom.ac.mu

DATES:
Received: 12 Oct. 2019
Revised: 07 Jan. 2021
Accepted: 19 Feb. 2021
Published: 29 Sep. 2021

HOW TO CITE:
Seebocus RH, Lollchund MR, Bessafi M. Analysis of extreme rainfall and drought events using statistical and fractal methods: A case study of Mauritius. S Afr J Sci. 2021;117(9/10), Art. #7477. https://doi.org/10.17159/sajs.2021/7477

ARTICLE INCLUDES:
☒ Peer review
☐ Supplementary material

DATA AVAILABILITY:
☐ Open data set
☐ All data included
☒ On request from author(s)
☐ Not available
☐ Not applicable

EDITOR:
Yali Woyessa

KEYWORDS:
extreme events, climate change, generalised extreme value distribution, Standardised Precipitation Index, fractal method

FUNDING:
None

Analysis of extreme rainfall and drought events using statistical and fractal methods: A case study of Mauritius

Due to climate change, extreme rainfall and drought events are becoming more and more frequent in several regions of the globe. We investigated the suitability of employing statistical and fractal (or scaling) methods to characterise extreme precipitation and drought events. The case of the island of Mauritius was considered, for which monthly mean rainfall data for the period January 1950 to December 2016 were analysed. The generalised extreme value distribution was used to extract the 10- and 20-year return levels and the Standardised Precipitation Index (SPI) was used to identify anomalous wet and dry events. A log-term correlation analysis was also performed to characterise the relationship between maximum rainfall and its duration. The results indicate that the 10-year return level is approximately between 500 mm and 850 mm and the 20-year return level is between 600 mm and 1000 mm. Results also show that the extreme maximum rainfall events occur mostly during austral summer (November to April) and could be related to the effects of tropical cyclones and La Niña events, while anomalous dry events were found to be significantly persistent with very long periods of drought. Moreover, there was a strong correlation between maximum rainfall and its duration. The methodology used in this work could be very useful in similar studies for other Small Island Developing States.

Significance:

- We show the usefulness of both statistical and fractal methods to understand occurrences of extreme precipitation events.
- We identify anomalous wet and dry events in rainfall time-series data using the Standardised Precipitation Index.

Introduction

Recent events occurring worldwide, such as the global heat record in 2016, major flood events in different parts of the world, the rapid melting of glaciers and sea level rise, among others, are reminders that the climate is changing more rapidly than in preceding centuries. The warming of our atmosphere and its effects on our natural and ecological systems are unavoidable and already palpable. Among the various natural disasters which affect humankind, heavy rainfall and long periods of drought have been reported to be detrimental to the environment, with disastrous consequences on the ecosystem, agriculture, and infrastructure as well as on the people facing such extremes.[1] For instance, in 2012, people in the central and western USA had to live in abnormally dry conditions due to persistent heat waves. This calamity put the health and safety of these people at risk and caused more than USD30 billion in damages.[2] Heavy rainfall can contribute to flash floods, resulting in water accumulation in flood prone areas, which is very dangerous. The year 2018 marked the occurrence of several deadly flash floods such as those in India, France and Vietnam, which caused several casualties.[3,4]

Heavy rainfall and severe droughts are global phenomena that affect major parts of the world, and are even seen to impact regions like Small Island Developing States.[5,6] Small Island Developing States are found mostly within the tropical belt and are very vulnerable to these calamities which are becoming increasingly frequent.[7] Moreover, with the growing populations on these islands and the increasing demand for a higher standard of living, it is important that decision-makers and urban planners have knowledge of extreme weather occurrences for better water resources management. To our knowledge, no rigorous study has been undertaken to quantify the statistical distributions of these calamities for Small Island Developing States. Results of statistical analyses of extreme events are necessary to help develop methods and strategies for flood disaster mitigation. We, therefore, aimed to reduce this void by employing statistical and fractal (or scaling) methods to characterise extreme precipitation and drought events. Our focus was on the island of Mauritius, and this study can serve as an example for similar islands.

Mauritius is a small island of volcanic origin and has a highly complex topography with an elevated central plateau. The geographical positioning of the island in the South West Indian Ocean region makes it vulnerable to diverse meteorological phenomena. Its prevailing weather is, however, mainly dominated by the monsoon regime and trade winds. The island is also subjected to the recurrent formation of tropical cyclones, thunderstorms and lightning as well as heavy rainfall in the summer seasons.[8-11] Mauritius' history is marked by unprecedented flash flood events which occurred in March 2008, February and March 2013, January 2015, February 2016, January 2018, December 2018 and, more recently, in January and February 2019. Some of these torrential rainfall events unfortunately caused the loss of human life and major property damage. The Mauritius Meteorological Services (MMS) characterise torrential rain conditions over the island as 'a prevailing weather which generates 100 mm of widespread rainfall in less than 12 hours and when such rainy conditions are likely to persist for several hours'[12].

Another hazard in the aftermath of heavy rainfall is landslides, which can occur on hill or mountain slopes which are considered high-risk areas. The MMS has pointed out that, during recent years, heat waves are lasting longer and the intensification trends of cyclones are steeper.[12] Hence, the impacts of climate variability in the form of extreme drought and rainfall events have become an issue of great concern due to the increasing dependence on rain-fed agricultural activities and the problem of water accumulation in some regions. Dhurmea et al.[13] presented a multi-temporal drought climatology for Mauritius for the period 1953–2007 using the Standard Precipitation Index (SPI). They produced high-resolution maps showing regions more prone to drought occurrence as well as regions with a rainfall surplus. Seebocus et al.[11] studied the statistical distribution of the rainfall data for Mauritius for the period 1950–2016 as well as the trend in the data using the Ensemble Empirical Mode Decomposition method. Their results indicate that three distributions – the log-normal, inverse Gaussian and the generalised extreme value (GEV) distribution – can be used to study the rainfall trend for Mauritius. The Ensemble Empirical Mode Decomposition analysis revealed that rainfall is strongly correlated to different meteorological phenomena such as tropical cyclones, quasi-biennial oscillations and the El-Niño Southern Oscillation (ENSO).

Raja and Aydin[10] analysed the rainfall variation of Mauritius by using data from 53 meteorological stations over a period of 30 years. The non-parametric Mann–Kendall and Spearman's rho statistical tests were used in the study to characterise the variation in annual precipitation. The results indicated both increasing and decreasing trends in the precipitation pattern across the island. It was noted that the majority of regions experienced a decrease in rainfall in the time interval 1996 to 2000, which was attributed to the 1998–2000 drought event contributed by the influence of a moderate La Niña event. Overall, an increase in the precipitation pattern was observed throughout the island. Raja and Aydin[14] extended their work to determine the regional rainfall trend of Mauritius using data collected at 52 meteorological stations for the period 1981 to 2010. Principal cluster analysis, cluster analysis and Kriging techniques were used to characterise the rainfall trend over the island. Their results indicated a decrease in rainfall between the southeast and northwest, which was attributed to the southeast trade winds and the altitudinal difference between the coastal areas and the central plateau.

Understanding the characteristics of heavy rainfall and drought at both temporal and spatial scales and estimating the return period associated with their occurrence is crucial in flood monitoring processes as well as in preparedness for severe drought events. It is also important to know the inter-relationship between these events and large-scale ocean–atmospheric processes.

Methodology

The study area and its climatic conditions

The island of Mauritius (Figure 1) is situated at latitude -20.2 and longitude 57.3. The island is influenced by tropical weather with two seasons, namely austral summer (November to late April) and austral winter (June to September), with May and October being transition months. The wet season extends from November to April (austral summer) with a farthest south position of Inter-Tropical Convergence Zone (ITCZ) resulting in strong convective rainfall activity. Conversely, from June to September (austral winter), active trade winds occur over Mauritius with more stable atmospheric air and less rainfall.[12,15] In addition, the precipitation is modulated on an inter-annual time scale (greater than one year) in relationship with the influence of large-scale circulation patterns such as the El-Niño Southern Oscillation, Indian Ocean Dipole and tropical cyclones.

According to the MMS, studies have confirmed an increasing trend in the occurrence of extreme weather and climate events in the South West Indian Ocean region.[16] The increase in temperature has impacted the hydrological cycle over the island. Long-term time series of rainfall amounts over the past century show a decreasing trend in annual rainfall over Mauritius; however, over the last two decades, there has been a significant increase in the frequency of severe weather events such as extreme precipitation causing flash floods due to water accumulation in flood prone areas and the formation of more intense tropical cyclones.[12]

Material and methods

Monthly mean rainfall data (CRU TS 4.01 data sets) for the period 1950–2016 obtained from the Centre for Environmental Data Analysis[17] were used for the analysis. This data set is a gridded satellite data product, on high-resolution (0.5×0.5 degree) grids, which is publicly available, and covers the period 1901–2016. The data includes climatic variables such as cloud cover, diurnal temperature range, frost day frequency, precipitation, daily mean temperature, monthly average daily maximum temperature, water vapour and wet day frequency. The data were checked for homogeneity using the Pettit test[18] and it was found that all data values before 1950 were redundant and were therefore omitted. Figure 2 shows the yearly rainfall pattern over Mauritius during the 1950–2016 period. It indicates the quasi-periodic behaviour of the rainfall distribution over the island. Figure 2 also includes occurrence of tropical cyclones during the period of study and the amount of precipitation received under the influence of the tropical cyclones. Table 1 highlights some characteristics of the data used for the study.

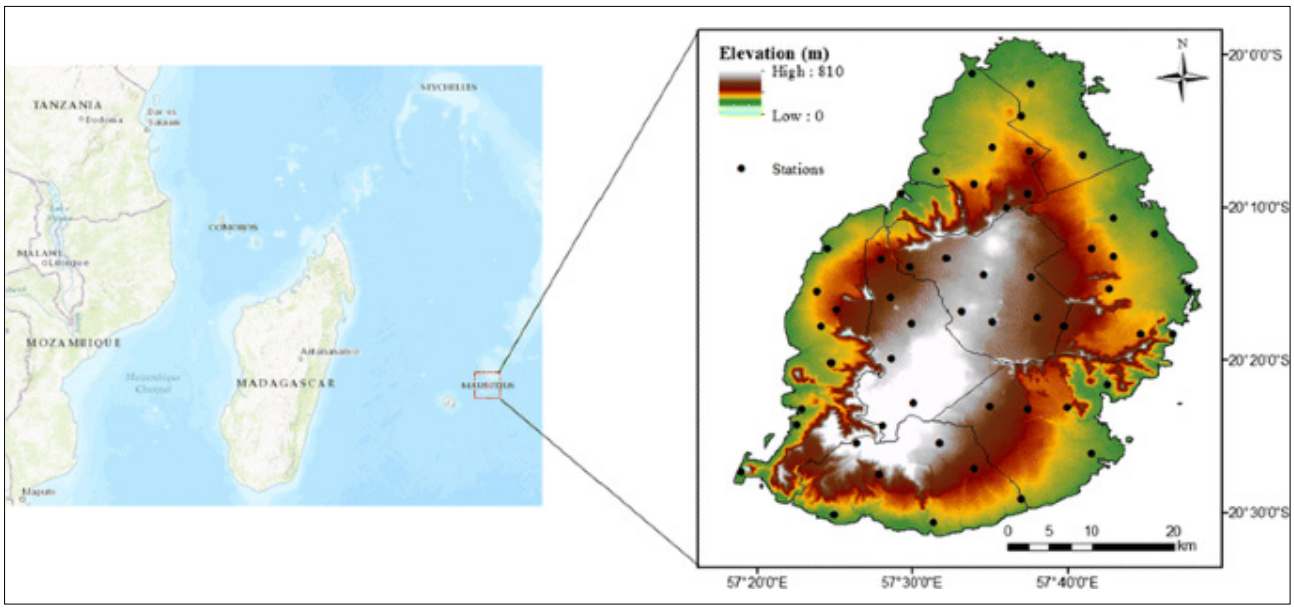


Figure 1: Location of Mauritius in the South West Indian Ocean region (left) and its topography (right). The different weather stations over the island are also shown.[14]

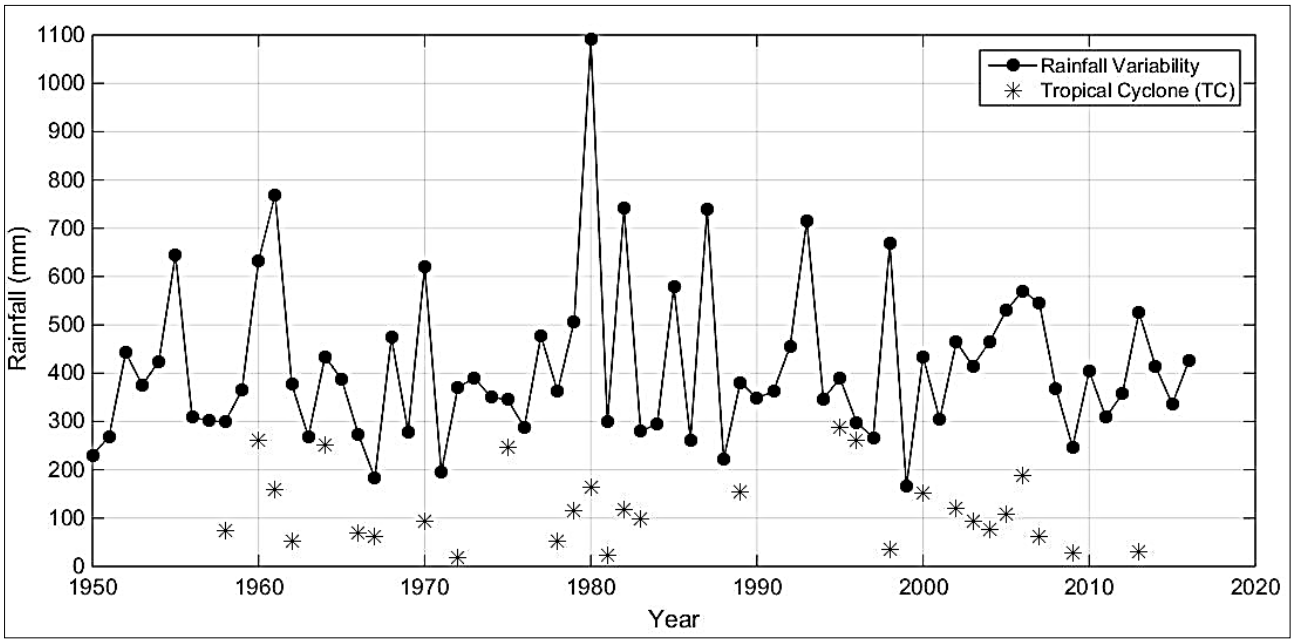


Figure 2: Yearly range (maximum minus minimum) of mean monthly rainfall over Mauritius during the 1950–2016 period. The asterisks indicate the occurrence of tropical cyclones.

Table 1: Characteristics of data used in the study

Statistic	Estimate
Number of records (*n*)	804
Minimum value (mm)	10.6
Maximum value (mm)	1126.5
Mean (mm)	164.8
Standard deviation (mm)	134.6
Skewness (mm)	2.03
Kurtosis (mm)	9.22

Extreme value analysis

Extreme value theory is the study of extreme and rare events based on the Fisher–Tipett theorem, which can be regrouped into two categories: the block maxima and the peaks over threshold. The block maxima method consists of distributing the data set into equal non-overlapping periods and determining the maximum that occurs in each period, thus restricting the analysis to only maximum values. The peaks-over-threshold method defines a threshold value and deals with all values that exceed the threshold.[19,20] As explained in the work of Ferreira and De Haan[20], the block maxima method is more efficient than the peaks-over-threshold method as it produces the least associated mean square error and smallest difference between the extreme value index and the quantile estimator. Seebocus et al.[11] showed that the frequency distribution of rainfall data for Mauritius (see Figure 3) can be represented using the GEV distribution. In this study, the block maxima method is used with the GEV distribution, where the cumulative distribution function is formulated as[21]:

$$F(x; \mu, \sigma, k) = \exp\left(- \left[1 + k \frac{(x - \mu)}{\sigma}\right]^{-1/k} \right) \quad \text{Equation 1}$$

where μ is the mean, σ is the standard deviation and k is the shape parameter. The aim of fitting the GEV distribution to the extreme rainfall data is to determine the return level of extreme events for specific return periods. The return level provides an estimate of the maximum rainfall that might occur at a specific time.[21]

The calculation procedure of the block maxima method and the return level estimation using the GEV distribution are summarised as follows[22]:

Assume the data series $X_1, X_2, \ldots \ldots, X_n$) is independent and identically distributed.

The maximum data point M_n obtained from the data series is assumed to be the data corresponding to the extreme event, and it is represented as

$$M_n = \text{Max}\ (X_1, X_2, \ldots \ldots, X_n) \quad \text{Equation 2}$$

Using the block maxima method, the cumulative distribution function can thus be written as indicated by Equation 3, where $F(x)$ represents the GEV cumulative distribution function.

$$P(M_n \geq x) = [F(x)]^n \quad \text{Equation 3}$$

The GEV distribution is fitted to the maximum data set M_n obtained and the GEV parameters (μ, σ, k) and the confidence intervals are estimated using the maximum likelihood estimation technique. The maximum likelihood estimation method is preferred as it is relatively simple to use compared to the Bayesian technique and it can be used to estimate the parameters of large data sets and their associated confidence intervals with little uncertainty.[23]

The return level occurring at a specific period can be estimated through interpolation of the results obtained.

Drought and flood analysis

Drought is a natural process, which occurs due to the lack of moisture in the atmosphere, which in turn results in a deficiency in precipitation for a prolonged time and this evolves slowly with time.[24] The effect of drought is worsened by other external factors such as high temperature, high wind speed and low moisture content. The impacts and duration of drought events are expected to increase due to the exacerbating effect of climate change.[25]

As identified by Wilhite and Glantz[26], drought events can be categorised into six types: meteorological, climatological, atmospheric, agricultural, hydrologic and water management. Drought behaviour is not the same for different water resources and keeps varying at different timescales, which make its analysis difficult.

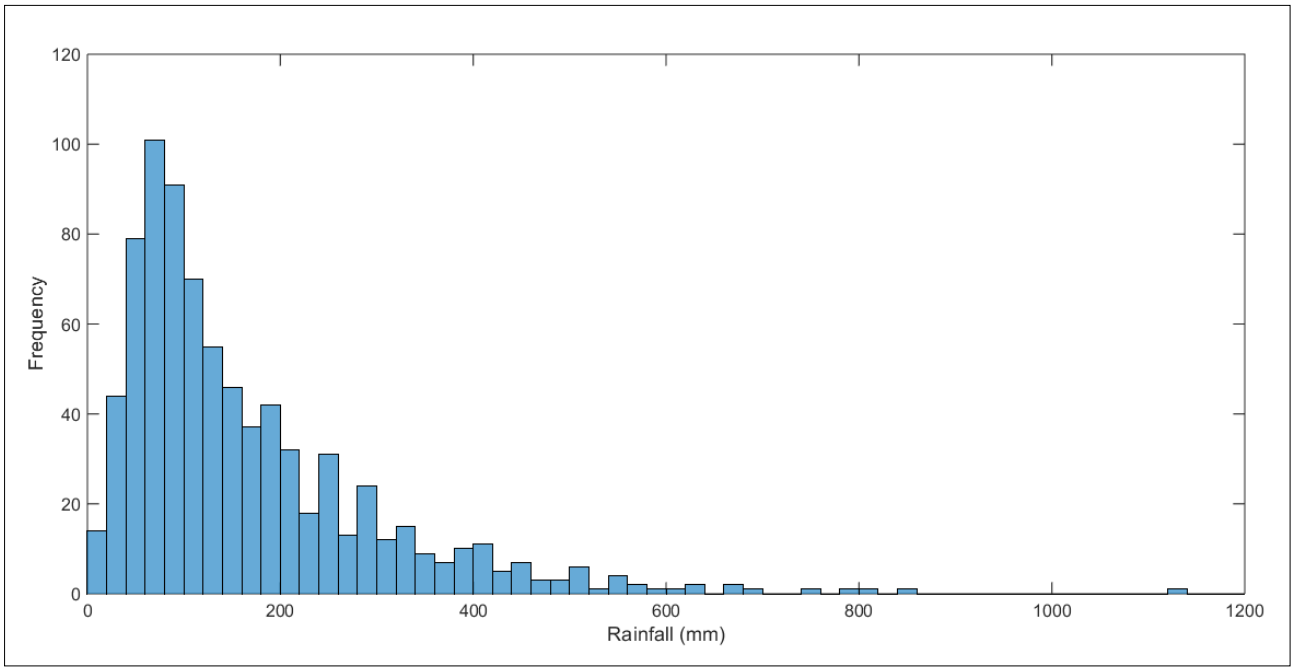


Figure 3: Monthly rainfall over Mauritius exhibits a skewed distribution with a heavy tail.

The SPI was proposed by McKee et al.[24] as the best drought monitoring index as it can be used to monitor all stages of a drought event from its onset to its completion for different water resources at different timescales. The SPI is a relatively simple tool which requires only precipitation data for a period of at least 30 years as the input parameter.

The SPI is calculated as follows[24,27]:

Fit the rainfall data to the Gamma probability distribution function which is given as:

$$g(x) = \frac{1}{\beta^{\alpha}\gamma(\alpha)} x^{\alpha-1} \exp\left(-\frac{x}{\beta}\right) \text{ for } x > 0 \qquad \text{Equation 4}$$

where α is the shape parameter, β is the scale parameter, x is the precipitation amount and $\gamma(\alpha)$ is the Gamma function defined by $\gamma(\alpha) = \int_{o}^{\infty} x^{\alpha-1} \exp(x)\, dx$.

Determine the timescale of i periods, where ($i = 1,2,3,\ldots..,72$ months) which depicts the number of months for which the drought will be investigated.

Estimate the corresponding parameters of the Gamma probability distribution function using the maximum likelihood estimation method.

Transform the result obtained to the standardised normal distribution to estimate the Z-distribution. This normalisation process is done as rainfall follows a skewed distribution as shown in Figure 3.

A drought event, as explained by McKee et al.[24], occurs when the SPI is continuously negative. A value less than -2 indicates extremely dry conditions. When the SPI becomes positive it indicates the end of dry conditions, and when it becomes greater than 1, wet conditions begin. Therefore, as the SPI is standardised, it can be used to monitor both dry as well as wet events at any timescales and locations and it is unaffected by topology.[25,28]

Scale and log-term correlation analysis

Rainfall data exhibit both spatial and temporal variations, which can be studied using fractal (or scaling) theory. The fractal patterns of rainfall data vary differently in both cases; at spatial resolution, the pattern is affected by geography of the study areas, while at temporal resolution the pattern is affected mainly by climatic conditions.[29] Jennings[30] was the first to propose a scaling law between the global maximum of rainfall and duration. The exponent of the power law discovered by Jennings was 0.5 for rainfall durations between 1 min and 1 year.

Rescaled range analysis (R/S), originally introduced by Hurst[31] is useful to measure the persistence and autocorrelation in time series. Such analysis assesses the Hurst exponent (H) of the power relationship between R/S parameter and duration. The Hurst exponent (H) is used as a measure of long-range dependence in time series.[32] As stated by Ceballos and Largo[32], there are different ways to determine the Hurst exponent, namely: adjusted rescale range analysis, detrended fluctuation analysis, and variance time plot analysis.

Discussions about these methods are beyond the scope of this work. The interested reader is referred to Ceballos and Largo[32]. These three methods have been tested and the results obtained confirmed that the adjusted rescale range analysis returned the lowest mean square error for all the estimated parameters. Hence it was employed in this work.

The Hurst exponent (H) is a statistical method commonly used to classify time series.[33] It has been used in diverse application areas such as biomedical signal processing, the financial sector, and climate change.[34] The Hurst exponent ranges between 0 and 1 and is classified into three categories as defined by Qian and Rasheed[33] and Valle et al.[29]:

$0 \leq H < 0.5$ (Indicates an anti-persistent time series)

$0.5 < H \leq 1$ (Indicates a persistent time series)

$H = 0.5$ (Indicates a random time series)

The first two categories represent the fractal Brownian movement while the third one represents the white noise or classic Brownian movement. A Hurst exponent of 0.5 indicates that there is no dependence between past and future values of the rainfall data under analysis and these types of data series are classified as unpredictable. It should be noted that the larger the value of H, the stronger will be the trend of the time series which increases as H approaches the value of 1.0.[29] The estimation of the Hurst exponent (H) is based on the work of Weron[35], whereby two different maximum likelihood estimation methods based on the peaks over threshold and the Block Maxima were used.[32]

Results and discussion

Return level estimation using the GEV distribution

As mentioned in the methodology section, the GEV distribution was used to determine the 10- and 20-year return levels and the block maxima method was used to extract the maximum rainfall data for each year during the period January 1950 to December 2016.

The maximum likelihood estimation equations used to estimate the parameters of the GEV distribution are given by Equations 5–8[11,36]:

$$\frac{1}{k}\sum_{i=1}^{n}\left[\frac{1-\sigma-(y_i)^{\frac{1}{\sigma}}}{y_i}\right]=0 \quad \text{Equation 5}$$

$$-\frac{n}{k}+\frac{1}{k}\sum_{i=1}^{n}\left[\frac{1-\sigma-(y_i)^{\frac{1}{\sigma}}}{y_i}\left(\frac{x_i-\mu}{y_i}\right)\right]=0 \quad \text{Equation 6}$$

$$-\frac{1}{\sigma^2}\sum_{i=1}^{n}\left\{\ln(y_i)\left[1-\sigma-(y_i)^{\frac{1}{\sigma}}\right]+\frac{1-\sigma-(y_i)^{\frac{1}{\sigma}}}{y_i}\sigma\left(\frac{x_i-\mu}{k}\right)\right\}=0 \quad \text{Equation 7}$$

$$y_i=[1-(\frac{\sigma}{k})(x-\mu)] \quad \text{Equation 8}$$

Table 2 summarises the result of the GEV parameters and the lower and upper bounds of the confidence intervals estimated using the maximum likelihood estimation method. The graph of return period against return level is shown in Figure 4. The 10- and 20-year return levels were interpolated from Figure 4; the 10-year return level was approximated to be between 500 mm and 850 mm and the 20-year return level was approximated to be between 600 mm and 1000 mm.

Table 2: Generalised extreme value (GEV) parameters estimated using the maximum likelihood estimation method

GEV Parameter	Confidence interval	
$\mu = 96.1227$	$\mu^- = 90.6817$	$\mu^+ = 101.5637$
$\sigma = 67.4783$	$\sigma^- = 62.8824$	$\sigma^+ = 72.4102$
k = 0.3529	$k^- = 0.2826$	$k^+ = 0.4233$

The temporal variation in monthly rainfall and the 10- and 20-year return levels are depicted in Figure 5. The occurrence of extreme wet events was mainly due to the influence on the local weather of tropical cyclones – Alix (January 1960), Beryl (December 1961), Louise (March 1970), Yacinthe (January 1980), Gabrielle (February 1982), Anacelle (February 1998), Hennie (March 2005) and Diwa (March 2006) and three strong El Niño events in the years 1982/1983, 1987/1988 and 1997/1998. The results show that the extreme maximum rainfall events are more prone to occur during austral summer seasons (November–April) as it is during this season that there are more cyclonic activities in the South West Indian Ocean region.

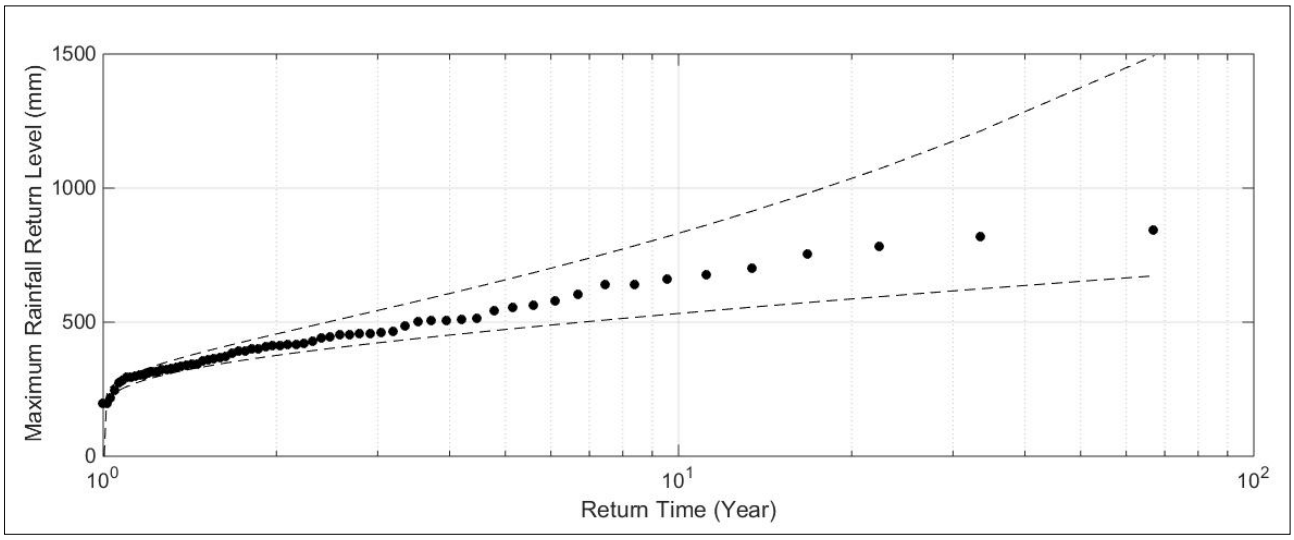


Figure 4: Yearly return level under generalised extreme value distribution over Mauritius during the period 1950–2016.

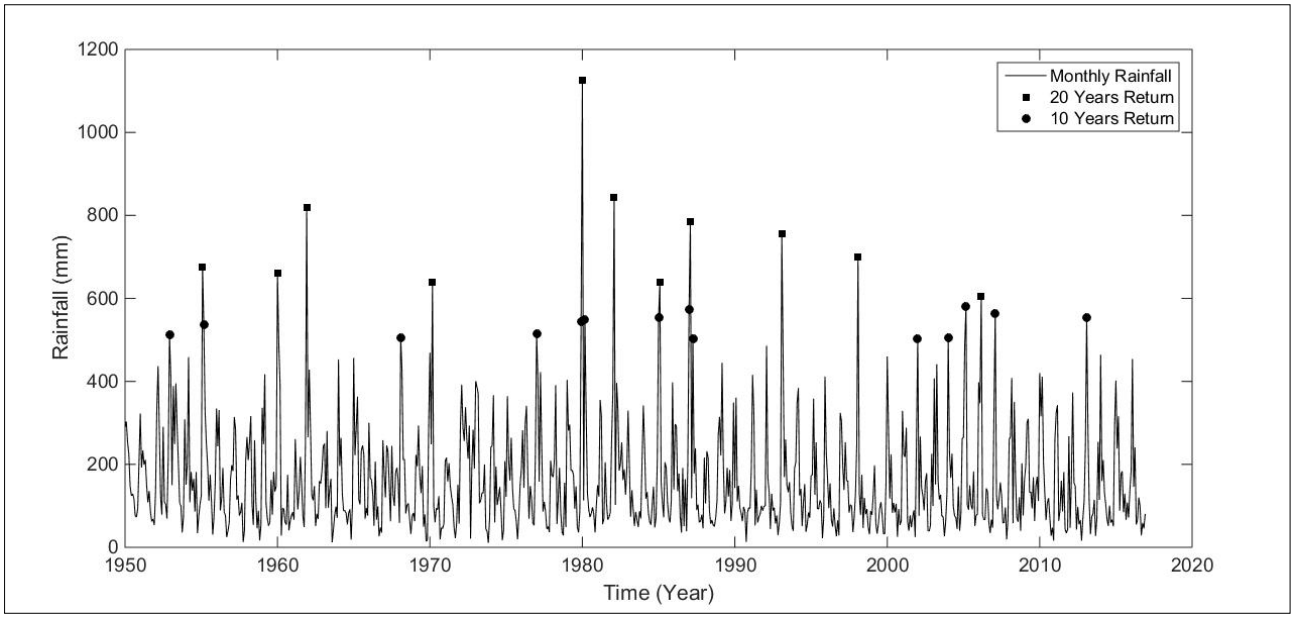


Figure 5: Monthly rainfall distribution and extreme events under 10- and 20-year return times.

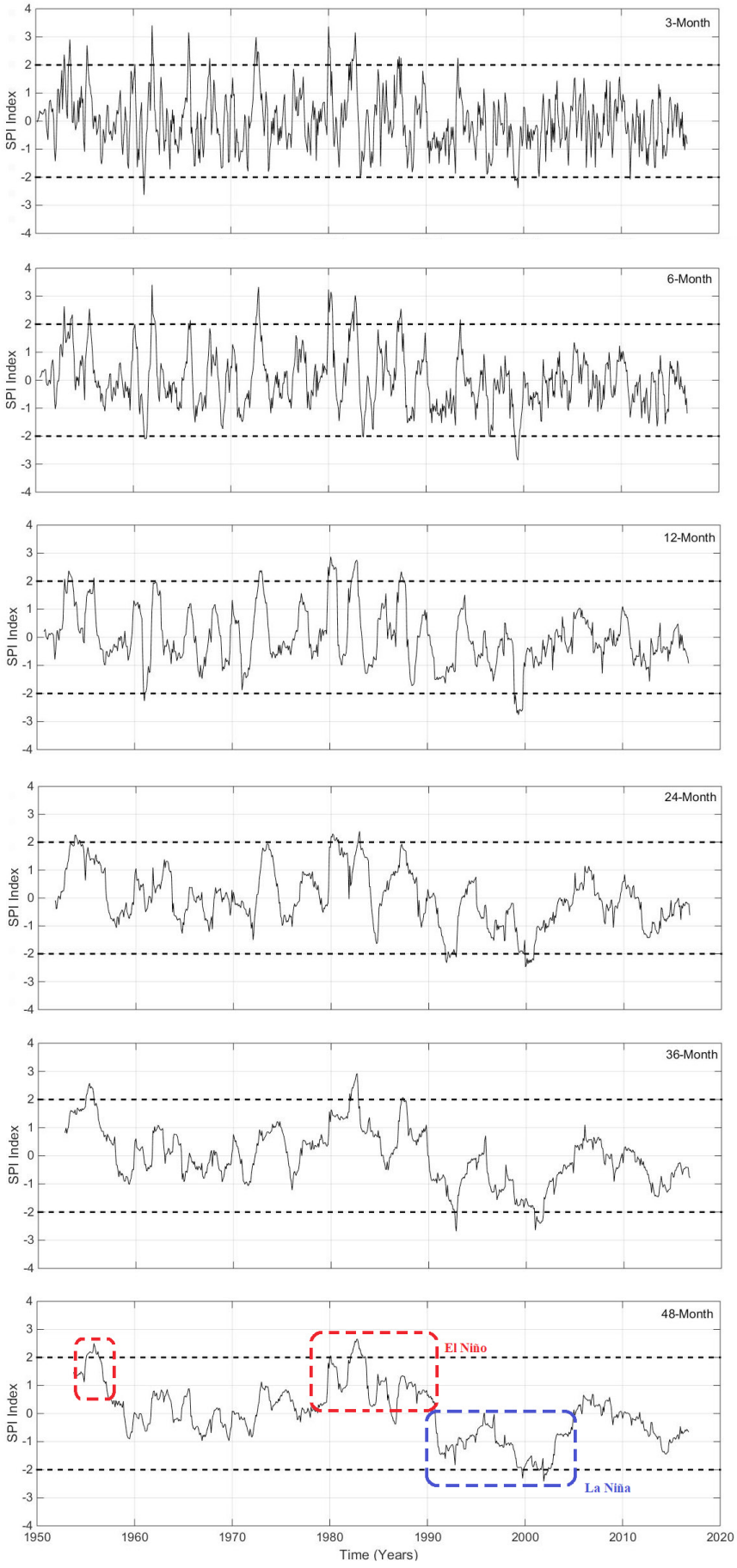


Figure 6: Standardised Precipitation Indices for rainfall over Mauritius during the period 1950–2016.

Drought analysis using the SPI

The SPI was estimated at several timescales (3, 6, 12, 24, 36 and 48 months) as shown in Figure 6. Drought events are characterised by negative values on the graphs; as soon as the line goes below zero, it marks the start of a drought event and the end of the drought is marked as soon as the line goes above zero again.[24,37] The results indicate that the 1995–1999 period was the longest period of drought for Mauritius (SPI< -2.0). Such an extremely dry event is noticeable on the overall timescale and could be considered as a significantly persistent event. The dry conditions could be related to the 1994/1995 and 1997/1998 La Niña events. For shorter timescales (3 and 6 months), extremely wet events could be related to cyclonic activities and El Niño events. The results obtained indicate that for the 36- and 48-month time scales, the frequencies of the dry events and wet events are almost the same, but the occurrence of dry events is higher than the occurrence of wet events, as indicated by the SPI graphs.

As explained in the work of Bordi et al.[38], the occurrence of dry conditions could be explained as a result of the influence of parameters such as orographic effect, land coverage, marine influence and altitude on the rainfall variability over the island.

Scale and log-term correlation

The scaling phenomena for the rainfall data is displayed in Figure 7. Superimposed over Figure 7 is the plot of the world record rainfall measurements of Zhang[39]. The best-fit lines indicate the relationship between maximum rainfall and duration. For the case of Mauritius, within a range of 1 month to 1 year, the maximum precipitation duration law exponent is 0.44 as compared to 0.9 (which indicates a random process with persistence) or the long-time range duration (>1 year).

Figure 8 displays the power law relationship between the *R/S* parameter and duration (log*(R/S)* versus log*(d)*). The Hurst exponent (H=0.6) derived from the graph indicates that the time series was found in the persistent regime range, as H was found in the range $0.5 < H \leq 1$. As stated in the literature, time series having larger H values can be predicted more accurately than those having an H value close to 0.5. The result obtained confirms that the Hurst exponent provides a good measure of predictability.

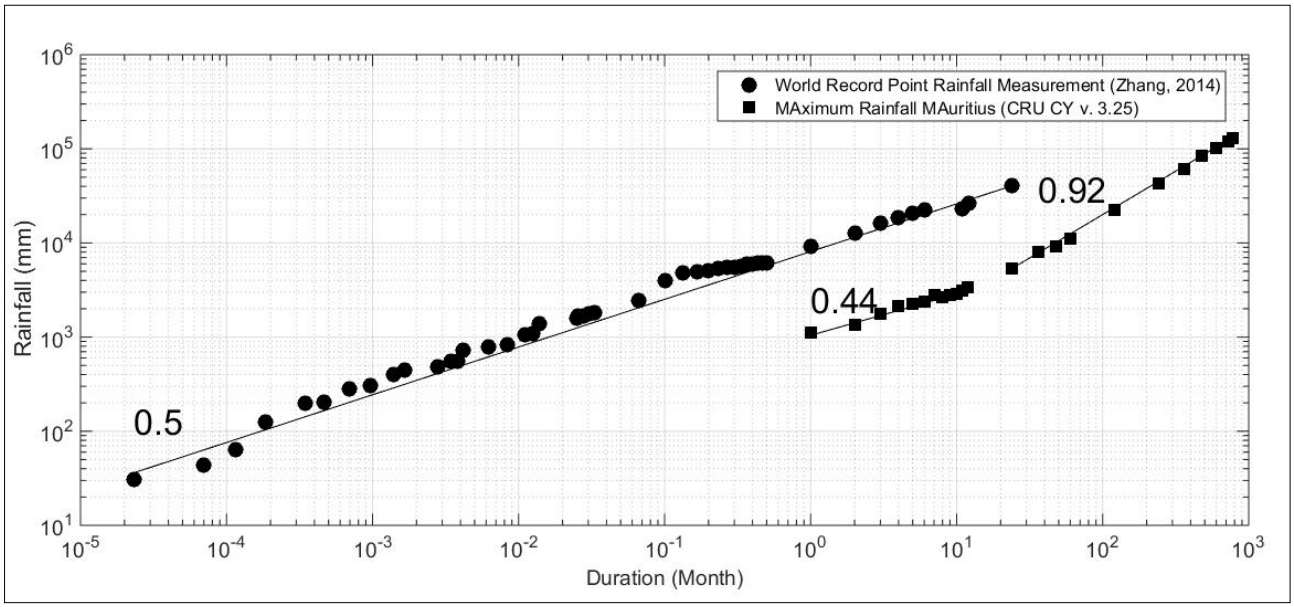


Figure 7: Rainfall-duration law scale observed over Mauritius during the period 1950–2016.

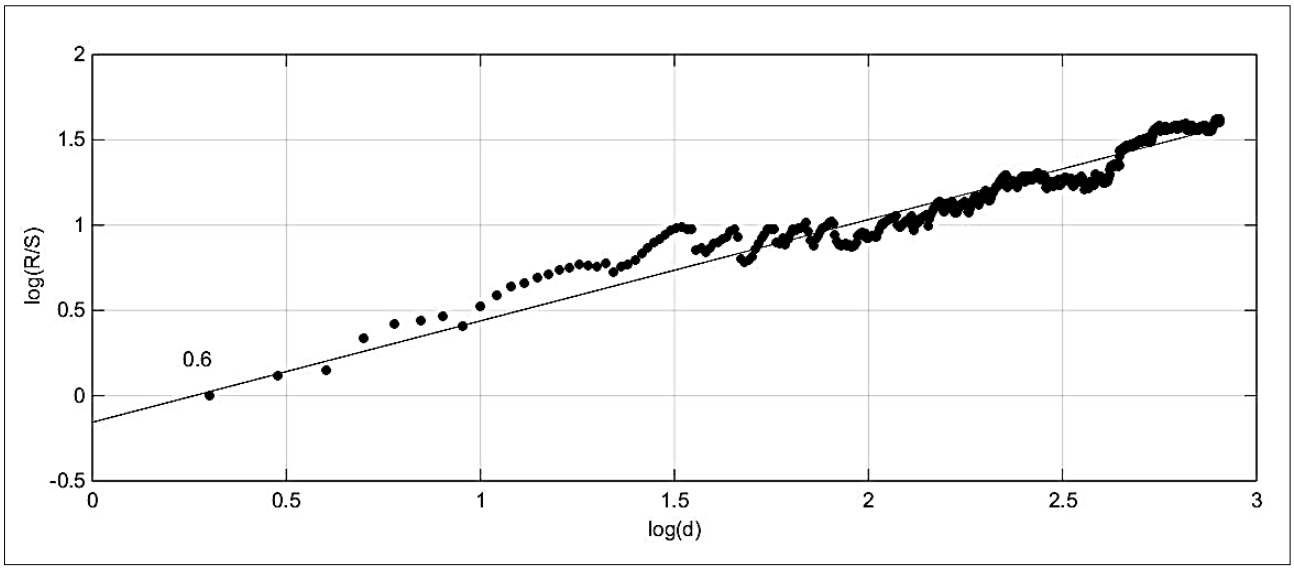


Figure 8: R/S versus duration to determine the Hurst exponent.

Conclusion

We have presented the analysis of return levels in extreme rainfall events occurring in Mauritius during the period 1950–2016. The GEV distribution was used to determine the 10- and 20-year return levels of extreme events. The GEV distribution was selected as the statistical tool for the estimation of the return level following the work of Seebocus et al.[11] who demonstrated that the GEV distribution best fitted the rainfall data for Mauritius, and could thus be used for statistical analysis.

The SPI was also employed to identify anomalous wet or dry events and to characterise the relationship between maximum rainfall and its duration. It was found that extreme maximum rainfall events occurred mostly during austral summer (November to April) and could be related to tropical cyclones and associated with dry conditions in the study region. The results identified the severe drought events which occurred from 1998 to 1990 due to the influence of the La Niña event. The SPI trends demonstrated that drier conditions occur more frequently than wet events over Mauritius. In addition, anomalous dry events were found to be significantly persistent with very long periods of drought and there was a strong correlation between maximum rainfall and its duration. Our findings demonstrate the usefulness of using both statistical and fractal methods to understand occurrences of extreme precipitation events.

Competing interests

We have no competing interests to declare.

Authors' contributions

R.H.S.: Data collection, sample analysis, data analysis, data curation, writing – the initial draft, writing – revisions. M.R.L.: Conceptualisation, writing – the initial draft, student supervision, project leadership, project management. M.B.: Methodology, student supervision, project leadership.

References

1. Climate News Network. Flash floods increase as mercury climbs. Flood List. 2018 Nov 13. Available from: http://floodlist.com/flood-risk/flash-floods-increase-as-mercury-climbs
2. Yin J, Gentine P, Zhou S, Sullivan SC, Wang R, Zhang Y, et al. Large increase in global storm runoff extremes driven by climate and anthropogenic changes. Nat Commun. 2018;9(1):4389. https://doi.org/10.1038/s41467-018-06765-2
3. Heavy storms cause deadly floods in south-western France. The Guardian. 2018 Oct 15. Available from: https://www.theguardian.com/world/2018/oct/15/heavy-storms-cause-deadly-floods-in-south-western-france
4. Taylor A. Devastating monsoon floods in Kerala, India. The Atlantic. 2018 Aug 22. Available from: https://www.theatlantic.com/photo/2018/08/devastating-monsoon-floods-in-kerala-india/568171/
5. Allam Z, Jones D. Towards a circular economy: A case study of waste conversion into housing units in Cotonou, Benin. Urban Sci. 2018;2(4):118. https://doi.org/10.3390/urbansci2040118
6. Allam Z. A theoretical application of the Extended Metabolism Model in Port Louis in a bid to promote urban sustainability. Paper presented at: 2nd International Conference on Energy, Environment and Climate Change; 2017 July 5–7; Mauritius.
7. Khoi DN, Trang HT. Analysis of changes in precipitation and extremes events in Ho Chi Minh City, Vietnam. Procedia Eng. 2016;142:229–235. https://doi.org/10.1016/j.proeng.2016.02.036
8. Senapathi D, Underwood F, Black E, Nicoll MA, Norris K. Evidence for long-term regional changes in precipitation on the East Coast Mountains in Mauritius. Int J Climatol. 2010;30(8):1164–1177. https://doi.org/10.1002/joc.1953
9. Staub CG, Stevens FR, Waylen PR. The geography of rainfall in Mauritius: Modelling the relationship between annual and monthly rainfall and landscape characteristics on a small volcanic island. Appl Geogr. 2014;54:222–234. https://doi.org/10.1016/j.apgeog.2014.08.008
10. Raja NB, Aydin O. Trend analysis of annual precipitation of Mauritius for the period 1981-2010. Meteorol Atmos Phys. 2019;131(4):789-805. https://doi.org/10.1007/s00703-018-0604-7
11. Seebocus RH, Lollchund MR, Bessafi M. Statistical and trend analyses of long-term rainfall data: A case study for Mauritius. Int J Hydrol Sci Technol. 2020;10(5). https://doi.org/10.1504/IJHST.2020.109945
12. Mauritius Meteorological Services. Climate Services: Climate change [webpage on the Internet]. No date [cited 2019 Feb 19]. Available from: http://metservice.intnet.mu/climate-services/climate-change.php
13. Dhurmea K, Boojhawon R, Rughooputh S. A drought climatology for Mauritius using the standardized precipitation index. Hydrolog Sci J. 2019;64(2):227–240. https://doi.org/10.1080/02626667.2019.1570209
14. Raja NB, Aydin O. Regionalization of precipitation in Mauritius: A statistical approach. Meteorol Appl. 2019;26:711–719. https://doi.org/10.1002/met.1798
15. McSweeney C, New M, Lizcano G, Lu X. The UNDP climate change country profiles: Improving the accessibility of observed and projected climate information for studies of climate change in developing countries. Bull Am Meteorol Soc. 2010;91(2):157–166. https://doi.org/10.1175/2009BAMS2826.1
16. Mauritius Meteorological Services. Climate Services: Seasonal climate forecast [webpage on the Internet]. No date [cited 2019 Feb 19]. Available from: http://metservice.intnet.mu/climate-services/seasonal-climate-forecast.php
17. Centre for Environmental Data Analysis. The CEDA Archive [webpage on the Internet]. c2018 [cited 2019 Mar 12]. Available from: https://archive.ceda.ac.uk
18. Javari M. Trend and homogeneity analysis of precipitation in Iran. Climate. 2016;4(3):44. https://doi.org/10.3390/cli4030044
19. Dombry C. Existence and consistency of the maximum likelihood estimators for the extreme value index within the block maxima framework. Bernoulli. 2015;21(1):420–436. https://doi.org/10.3150/13-BEJ573
20. Ferreira A, De Haan L. On the block maxima method in extreme value theory: PWM estimators. Ann Stat. 2015;43(1):276–298. https://doi.org/10.1214/14-AOS1280
21. Boudrissa N, Cheraitia H, Halimi L. Modelling maximum daily yearly rainfall in northern Algeria using generalized extreme value distributions from 1936 to 2009. Meteorol Appl. 2017;24(1):114–119. https://doi.org/10.1002/met.1610
22. Shukla RK, Trivedi M, Kumar M. On the proficient use of GEV distribution: A case study of subtropical monsoon region in India. Ann Comput Sci Ser. 2010;8(1):81–93.
23. Coles S, Pericchi LR, Sisson S. A fully probabilistic approach to extreme rainfall modeling. J Hydrol. 2003;273(1–4):35–50. https://doi.org/10.1016/S0022-1694(02)00353-0
24. McKee TB, Doesken NJ, Kleist J, editors. The relationship of drought frequency and duration to time scales. In: Proceedings of the 8th Conference on Applied Climatology; 1993 January 17–22; Anaheim, CA, USA. Boston, MA: American Meteorological Society; 1993.
25. Baran T, Bacanli UG, Dikbas F. Drought analysis with SPI index and entropy. Eur Water. 2017;60:263–270.
26. Wilhite DA, Glantz MH. Understanding: the drought phenomenon: The role of definitions. Water Int. 1985;10(3):111–120. https://doi.org/10.1080/02508068508686328
27. Lloyd-Hughes B, Saunders MA. A drought climatology for Europe. Int J Climatol. 2002;22(13):1571–1592. https://doi.org/10.1002/joc.846
28. Portela M, Zeleňáková M, Santos J, Purcz P, Silva A, Hlavatá H. A comprehensive drought analysis in Slovakia using SPI. Eur Water. 2015;51:15–31.
29. Valle MV, García GM, Cohen IS, Klaudia Oleschko L, Ruiz Corral J, Korvin G. Spatial variability of the Hurst exponent for the daily scale rainfall series in the state of Zacatecas, Mexico. J Appl Meteorol Clim. 2013;52(12):2771–2780. https://doi.org/10.1175/JAMC-D-13-0136.1
30. Jennings AH. World's greatest observed point rainfalls. Mon Weather Rev. 1950;78(1):4–5. https://doi.org/10.1175/1520-0493(1950)078<0004:WGOPR>2.0.CO;2

31. Hurst HE. Long-term storage capacity of reservoirs. Trans Am Soc Civil Eng. 1951;116:770–799. https://doi.org/10.1061/TACEAT.0006518

32. Ceballos RF, Largo FF. On the estimation of the Hurst exponent using adjusted rescaled range analysis, detrended fluctuation analysis and variance time plot: A case of exponential distribution [preprint]. arXiv. 2018. arXiv:180508931. Available from: https://arxiv.org/abs/1805.08931

33. Qian B, Rasheed K. Hurst exponent and financial market predictability. Proceedings of IASTED Conference on Financial Engineering and Applications. 2004:203–209.

34. Kulish V, Horák V. Forecasting the behavior of fractal time series: Hurst exponent as a measure of predictability. Review of the Air Force Academy. 2016(2):61. https://doi.org/10.19062/1842-9238.2016.14.2.8

35. Weron R. DFA: MATLAB function to compute the Hurst exponent using Detrended Fluctuation Analysis (DFA). HSC Software M11002. Wroclaw: Hugo Steinhaus Center, Wroclaw University of Technology; 2011.

36. Martins ES, Stedinger JR. Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. Water Resour Res. 2000;36(3):737–744. https://doi.org/10.1029/1999WR900330

37. Hayes MJ, Svoboda MD, Wiihite DA, Vanyarkho OV. Monitoring the 1996 drought using the standardized precipitation index. Bull Am Meteorol Soc. 1999;80(3):429–438. https://doi.org/10.1175/1520-0477(1999)080<0429:MTDUTS>2.0.CO;2

38. Bordi I, Frigio S, Parenti P, Speranza A, Sutera A. The analysis of the Standardized Precipitation Index in the Mediterranean area: Large-scale patterns. Ann Geophys. 2001;44(5–6):965–978. https://doi.org/10.4401/ag-3549

39. Zhang H. Extreme rainfall events in simulations, theory and related large-scale dynamic processes [dissertation]. Hamburg: University of Hamburg; 2014.

AUTHORS:
Beatrice Conradie[1]
Jenifer Piesse[2,3]
Johann Strauss[4]

AFFILIATIONS:
[1]School of Economics, University of Cape Town, Cape Town, South Africa
[2]Department of Management, King's College London, London, United Kingdom
[3]Department of Agricultural Economics, Stellenbosch University, Stellenbosch, South Africa
[4]Western Cape Department of Agriculture, Stellenbosch, South Africa

CORRESPONDENCE TO:
Beatrice Conradie

EMAIL:
Beatrice.conradie@uct.ac.za

DATES:
Received: 16 Sep. 2020
Revised: 21 Apr. 2021
Accepted: 04 May 2021
Published: 29 Sep. 2021

HOW TO CITE:
Conradie B, Piesse J, Strauss J. Impact of heat and moisture stress on crop productivity: Evidence from the Langgewens Research Farm. S Afr J Sci. 2021;117(9/10), Art. #8898. https://doi.org/10.17159/sajs.2021/8898

ARTICLE INCLUDES:
☒ Peer review
☐ Supplementary material

DATA AVAILABILITY:
☐ Open data set
☐ All data included
☒ On request from author(s)
☐ Not available
☐ Not applicable

EDITORS:
Teresa Coutinho
Salmina Mokgehle

KEYWORDS:
total factor productivity, climate, rainfed arable farming, experimental farm environment, Langgewens Research Farm

FUNDING:
None

Impact of heat and moisture stress on crop productivity: Evidence from the Langgewens Research Farm

We investigated the effect of heat and moisture stress on total factor productivity in crop farming under experimental farm conditions. Heat stress is the number of days during the growing season during which the maximum temperature exceeds 24.9 °C. Total rainfall is treated as a basic factor of production and periodic moisture stress, or lack thereof, is the number of rainfall days during the growing season. All models controlled for the cumulative soil benefits arising from minimum tillage, which is the main objective of the experiment. Model specification was evaluated using likelihood ratio tests and three are worthy of note. The study site received 329 mm of rainfall on average on 22 rainy days per season during the period 2002–2015, while the maximum temperature typically rose above 24.9 °C on 33 days per growing season. The average efficiency of the plots in the long-term crop rotation experiment increased at 3.4% per year from a base of 60% to the most recent level of 78%. Neither heat nor moisture stress changed significantly over the study period. Heat stress was found to reduce efficiency by 1.75% per hot day and rainfall increased efficiency by 1.45% for each additional rainy day. However, the interaction of heat and moisture stress lowered productivity overall.

Significance:

- This study contributes a new approach to modelling the effect of climate on agricultural productivity using a new metric of heat and moisture stress.
- We quantify the marginal effects of rising temperatures and rainfall events and evaluate several potential specifications of heat and moisture stress variables.

Introduction

A global rise in temperature is a feature of climate change and is likely to impact rainfall, both in the level and in the distribution. However, there remains some uncertainty about the heat and moisture stress relationship between these environmental phenomena. It is predicted that 20% lower rainfall combined with a 2 °C increase in temperature would reduce profitability in field crops in South Africa by 4.4%.[1] That is, if the temperature rose by 3 °C, profits would fall by 11.7%, although should this increase be accompanied by moderate rainfall, real profits may increase. The prediction for the Northern and Western Cape Provinces of South Africa is a 1.5–2 °C rise in surface temperatures combined with a 5–10% lower median rainfall by the turn of the century.[2] Thus, if the predictions are right, there is no real threat overall from climate change, except for wheat production in the winter rainfall area, which will be stressed by rising temperatures.

However, given the absence of any certainty with respect to long-term meteorological forecasts, we investigated the effects of current rainfall and temperature ranges on the performance of dryland wheat production in the Western Cape. Clearly, the more precisely the climate variables can be measured, the more valuable the efficiency of producers facing these environmental factors is to policymakers and practitioners alike. Our approach was based on Ricardian climate models, although here there is a different dependent variable from those used earlier, as well as an emphasis on capturing climate stress. In the Mendelsohn et al.[3] analysis, land prices were used as a proxy for expectations about future income. By explaining land prices with average rainfall and temperature, cross-section variation can be used to predict how changing climate is likely to affect the global food system, given controls for soil potential etc. Due to the lack of suitable farm sales data, the World Bank abandoned land prices as a dependent variable in their studies in favour of net farm income or yield[4], and this approach is replicated here. Sales and net farm income are closely correlated, with expected net revenue per hectare obtained from data on farm sales prices by assuming a suitable discount rate, whereas yield is equivalent to net revenue at fixed prices. In both cases it is important to assume a given level of technical progress and expectations about how climate change will present in the future. If either yield or net revenue replaces land prices, the model loses much of its original elegance, as these additional factors now have to be controlled for explicitly.

Review of the literature and contribution of this paper

Ricardian models are frequently used to analyse agricultural production and are derived from the simple observation that the value of land reflects its net productivity. Most authors use a cross-sectional approach. A major influence on this paper was Mendelsohn et al.'s[3] study in which controls were included to account for eight soil characteristics, along with altitude, latitude (as a proxy for solar radiation), per capita income and population density, the latter to take account of opportunities in the non-farm sector. Yield was modelled as a quadratic function of rainfall and temperature whilst seasonal rainfall and temperature effects were included as separate variables. Gbetibouo and Hassan[1] build on Mendelsohn et al.[3] by using the Ricardian model to capture the effect of rainfall and temperature variability on land productivity and land values, although they only considered long-term spatial variation and ignored temporal variability between and within seasons. Kurukulasuriya et al.[5] modelled net farm income per

hectare in several African countries, using access to electricity as a proxy for modern infrastructure, while Gbetibouo and Hassan[1] introduced the size of the farm labour force in addition to population density to capture potential macroeconomic shifts.

The Ricardian approach has continued to evolve, such as moving to more interesting dimensions and transformations of the data, including panel data[6] and first differences in a time series model[7]. Cabas et al.[8] summarised many of the site characteristics in an area change variable based on the supposition that yields will fall and become more erratic when production expands onto marginal land. The same analysis included technical change and input price variation and introduced climate volatility using the mean of the coefficient of variation (the standard deviation divided by the mean) of both rainfall and temperature. Temperature is the mean daily value while rainfall is in total millimetres recorded during the growing season. Following the principles of phenology, the growing season begins when temperatures rise above 5 °C for five consecutive days and is computed from monthly average data. In Boubacar[9], the yield response model was reduced to two moisture stress variables, beneficial temperature was measured in growing degree days, technical progress and a somewhat dubious area change variable accounted for unexplained dynamics. The first moisture stress variable captured drought periods as a percentage of expected rainfall while the second identified the month with the highest share of annual rainfall to capture variability.

The literature on Ricardian modelling is clear about how to capture climate fixed effects; yield is usually specified as a quadratic function of total rainfall and mean temperature. Dinar et al.[10] went a step further by defining an aridity index as mean temperature divided by mean rainfall, which formed part of the frontier sub-model. Roux[11] expressed a similar idea by defining a relative drought tolerance index as rainfall squared divided by its standard deviation, although this index was never used in productivity modelling. With respect to estimation, it became important to consider seasonal differences as Ricardian models began to be estimated with panel data. Cabas et al.[8] captured variability as a fixed effect with the coefficient of variation of rainfall, temperature and growing degree days, although this method still does not reflect the difference in variability between one season and the next. Employing adverse climate as an explanation for the observed differences in farm-level efficiency requires a specification that captures climate stress, which is site specific. In the northern latitudes of Canada, temperature stress comes in the form of low temperatures, best measured as growing degree days[8], although this does not limit wheat growth in South Africa while heat stress does.[12]

This paper contributes to the literature in two ways. Firstly, there is currently no credible model that considers comprehensive weather changes in productivity. Therefore, we used time series data from a multi-plot crop rotation experiment at a single research site and focussed on measuring year-on-year differences in the amount of heat and moisture stress. Secondly, we used a novel approach to estimation, by extending the standard stochastic frontier production function with inefficiency effects.[13] In this model, a best-practice frontier was jointly estimated with an inefficiency model that explains individual performance relative to the benchmark. Inputs and outputs into the production process were considered and the typical explanations for inefficiency in this well-rehearsed literature usually include some combination of technology, subsidies, governance and extension factors. Resource quality, such as access to irrigation, has formed part of the explanation right from the start and can appear as part of the frontier or in the inefficiency model.[13,14] Finally, by adding enhanced proxies for moisture and heat stress, the standard model was adapted for monitoring climate change impacts on agricultural productivity.

Methods and data

The established dependent variable in the Ricardian model is yield, or profitability, where the latter is yield at current prices for a given technology. Alternatively, land values can be used, which is the present value of the expected future income stream at constant prices. We propose that a total factor productivity score replace yield as the dependent variable. This measure has the advantage of being independent of prices and capable of capturing both technical change and changing factors of production.

The stochastic frontier production function model with climate-based inefficiency effects uses input and output data to fit a benchmark (Equation 1) and various proxies for temperature and rainfall limitations to production to explain deviations from the benchmark (Equation 2). Instead of referring to farms, as is usually done, the inefficiency scores predicted by this two-part model refer to experimental plots of a quarter hectare each. These plots form part of one of four crop rotation systems that are being compared to other crop–livestock rotation systems at Langgewens Research Farm in the Western Cape, South Africa (33°17'0.78''S, 18°42'28.09"E). The data were provided by the researchers at the farm and we checked the information using standard robustness tests. The site's average annual rainfall since 1964 is 403 mm and almost 80% of it falls during the winter growing season, from April to September.

The productivity model can be stated as:

$$\ln Y_{it} = \alpha_0 + \sum_{k=1}^{K} \alpha_k \ln x_{kit} + v_{it} - u_{it} \qquad \text{Equation 1}$$

and

$$-u_{it} = \beta_0 + \sum_{m=1}^{M} \beta_m \cdot z_{mit} + w_{it} \qquad \text{Equation 2}$$

where Y is output, x is input, z potential explanations for deviations from the frontier (inefficiency effects), and α and β are parameters to be estimated. The error term w_{it} in Equation 2 is a typical normally distributed error term. In Equation 1, the error term is decomposed into a normally distributed component, v_{it}, and a one-sided inefficiency term, u_{it}, which captures each observation's degree of deviation from the benchmark. Output is measured as the natural logarithm of the real value of product sales and nominal values were deflated using the general consumer price index published in the Abstract of Agricultural Statistics.[15] The rotations incorporated here are a wheat monoculture, a wheat–canola rotation and two systems that rotate wheat and canola with lupins. Two thirds of the observations are for wheat, while 21% apply to canola and 14% to lupins. The data are from one crop rotation trial at the research facility. Decisions like fertiliser applications and planting dates are jointly controlled by the responsible researcher and farm manager. Inputs are bought on tender and output is sold on the open market. The inputs are seed and fertiliser cost, chemicals (pesticides herbicides and fungicides), mechanisation cost and total seasonal rainfall. Rainfall is measured in millimetres recorded during the growing season (April – September). The inputs in value terms are in constant 2010 ZAR prices deflated according to the input specific deflators in the Abstract and logged. As in all well-behaved production functions, α is expected to be positive and significant.

Pooled descriptive statistics for the study period, 2002–2015, are shown in Table 1. Land is obsolete as plot data are expressed per hectare, and, because labour is used in fixed proportion to machinery, it is omitted to avoid collinearity.

Instead of modelling crop performance as a function of mean rainfall and temperature as Ricardian models do, we specifically wanted to capture heat and moisture stress on the total factor productivity of each crop in the production system. The simplest formulation for heat stress is a count of growing days on which the maximum temperature reaches an arbitrary threshold. After experimenting with several, we opted for 24.9 °C which is often used for wheat.[12] While output is correlated with total rainfall, total rainfall does not capture the effect of periodic moisture stress. The standard deviation of rainfall, or its coefficient of variation, has been used as a measure of variability.[8] Given the construction of this statistic, the higher the standard deviation for a given level of rainfall, the higher its coefficient of variation. With total rainfall already in the stochastic frontier, it was logical to use the standard deviation of rainfall in the inefficiency model. This was calculated from daily observations over the growing season (April – September) and predicted a higher standard deviation to cause more inefficiency. The product of the two, which would capture the interaction of temperature and moisture stress, was predicted to increase inefficiency.

Table 1: Descriptive statistics for the pooled sample of 28 plots over 14 years (n=392)

Variable name	Description and units	Mean	s.d.
Output	Real income / ha	7620.64	4422.28
Seed and fertiliser	Real cost / ha	2818.01	806.25
Chemicals	Real cost / ha	1071.31	332.56
Mechanisation	Real cost / ha	786.70	205.88
Rainfall	mm, April – September	328.66	98.13
Heat stress	Days ≥ 24.9 °C, April – September	33	6.14
Rainy days	Days ≥ 5mm, April – September	21.71	5.83
Aridity index	Average over the growing season of monthly rainfall / average temperature in that month		

A number of specifications were estimated. Dinar et al.[10] combined temperature and heat effects into an aridity index, defined as the ratio of annual mean daily temperature over total annual rainfall. This variable was tested independently and in combination with the seasonal heat and moisture stress variables described above. The aridity index is for the growing season only. Mean daily temperature was calculated by taking the average of the daily minimum and maximum and then the average over the growing season to compute the average mean daily temperature for the growing season. This was divided by total rainfall recorded during the growing season. The prediction was that greater aridity would increase inefficiency.

In another specification, the standard deviation of rainfall as a proxy for moisture stress was replaced by a count of rainy days, defined as 24-h periods that receive more than 5 mm of rainfall. For a given total seasonal rainfall an increase in the number of rainy days implies shorter periodic droughts, which ought to decrease inefficiency. Generalised likelihood ratio tests were used to choose between nested specifications, but this test does not allow the choice between the different ways of capturing periodic drought. There we were guided by the overall goodness of fit and the signs and significance of the input elasticities.

This study took place against the backdrop of soil improvements after the adoption of zero tillage, which is expected to raise productivity. Langgewens Research Farm switched to minimum tillage in 1996 and adopted zero tillage in 2002. Dramatic improvements in soil conditions followed and this is captured in Equation 2 by a time trend, which is the usual method of accounting for technical change. Winter rains failed in 2003 (67% of expected seasonal rainfall), 2004 (62%) and 2015 (54%), and 2007 was an exceptionally good year (141% of expected seasonal rainfall). In 2003, most plots performed sufficiently well to be harvested. By 2015, the same plots did well under similar conditions. The soil gains include higher carbon levels that improve soil structure, permeability and fertility, and on some plots rotating monocotyledon with dicotyledon crops lowers weed pressures enough to cut down on herbicide inputs.[12]

As no new varieties or production methods featured in the experiment during the study period and given the low technical progress in dryland crop farming on commercial farms in the Western Cape during the second half of the 20th century[16], it was unlikely that the system would also experience Hicks-neutral technical progress and so there was no need to include a time trend in the frontier model.

$$TE_i = \frac{Y_i}{Y_i^*} = \frac{f(x_i;\beta)(v_i - u_i)}{f(x_i;\beta)(v_i)} \qquad \text{Equation 3}$$

Equation 4 gives the Battese and Corra[17] parameterisation of the inefficiency term, in which gamma is calculated as follows:

$$\gamma = \sigma_u^2 \Big/ \sigma_v^2 + \sigma_u^2 \qquad \text{Equation 4}$$

The efficiency scores predicted by Equation 3 vary from zero to one, or 0–100%. Observations close to 100% set the benchmark, but as the model allows for statistical noise, the best performing plots are usually no more than 97% efficient under optimal conditions, although mean scores vary with model complexity and sample size. For experimental plot data, where mismeasurement is negligible, gamma could approach unity.

The translog functional form has become standard in total factor productivity estimates, and is often accompanied by a log-likelihood test that compares its performance to that of Cobb Douglas. The benefit of estimating a more general functional form is that it relaxes the assumption of constant elasticities of substitution made by Cobb Douglas. However, in this case, in which there is a single manager in control of day-to-day production decisions for all plots on the experimental farm, the Cobb Douglas was considered sufficient, especially as it allows more degrees of freedom to experiment with climate variables.

Results and discussion

This section is divided into four parts. The first part presents the baseline model, which has only the cumulative no tillage benefits and heat stress in the inefficiency equation. The second part introduces periodic drought stress in the form of a standard deviation on seasonal rainfall as a third z-variable. The number of rainy days replaces the standard deviation of rainfall in the third part and the performance of the Dinar aridity index is evaluated in the fourth part. The same procedure was followed with all three proxies and all results in the tables first present the frontier specification and then the results of the inefficiency model. Once the suitability of the basic proxy had been confirmed, it was interacted with the heat stress variable to determine if there was a joint effect.

The baseline model with no tillage benefits and heat stress

The baseline specification works well here. All four Cobb Douglas input coefficients have the expected sign and are significant at $p \leq 0.10$. The output elasticities indicate that output is most closely correlated to total seasonal rainfall, where a 1% increase in rainfall will result in a 0.49% increase in the real value of output. The second largest output elasticity is on seed and fertiliser, where a 1% increase in expenditure is predicted to result in a 0.40% increase in output. This is followed by mechanisation whose output elasticity is 0.179 and agro-chemicals whose output elasticity is 0.076. The relative unimportance of agro-chemicals in this production system explains why its coefficient is measured with such a relatively low level of certainty.

The inefficiency model performs equally well. By relaxing the mean response function assumption of independently and identically distributed error terms, the stochastic frontier model presented in Table 2 is estimated with four additional parameters – sigma-squared, gamma,

the no tillage benefit and the heat stress coefficient, all of which are significantly different from zero. The log-likelihood value of the mean response function was -752.99, while the stochastic frontier model yielded a value of -377.59. A likelihood ratio test was performed to test that the four restrictions were valid. The test statistic of LR = -2(-752.99 – (-377.99)) = 750.80 rejects the mean response model at the highest level of significance. The critical value is 12.483 for $p \leq 0.01$. It is clear from these tests that the stochastic frontier function form fits the data better than the ordinary least squares estimation. The mean efficiency of the pooled sample is 55%, the cumulative benefit of the practice of zero tillage is 3.45% per year and an extra day of heat stress during the growing season reduces efficiency by 1.78% across all crops and rotations in the sample.

Table 2: Baseline grain productivity model (Model 0) with no till benefits and heat stress

Variable	Coefficient	s.e.	*t*-ratio	Significance
Constant	1.792	0.801	2.236	**
Seed and fertiliser	0.397	0.097	4.111	***
Chemicals	0.076	0.039	1.953	*
Mechanisation	0.179	0.083	2.157	**
Rainfall	0.492	0.045	10.893	***
Constant	-32.284	1.834	-17.600	***
No till benefit	-1.523	0.152	-9.992	***
Heat stress	0.925	0.049	18.954	***
Sigma-squared	10.659	0.896	11.895	***
Gamma	0.999	0.000	2809.206	***
Log-likelihood statistic	-377.59			
Observations	392			

***p ≤ *0.01* , **p ≤ *0.05* , *p ≤ *0.10*

Is the coefficient of variation of rainfall a valid proxy for periodic moisture stress?

Introducing the coefficient of variation of seasonal rainfall into the inefficiency sub-model produced a plausible stochastic frontier production function and confirmed the expectations about the effects of zero tillage and heat stress on plot-level efficiency. The output elasticities are all significant and have similar magnitudes as before. The no tillage benefit increases fractionally to 3.54% per year. The negative impact associated with an additional day of heat stress goes up to 1.79%. The restrictions imposed by the mean response function are rejected with a log-likelihood test statistic of 793.79, and when compared to the result in Table 2, the hypothesis that the standard deviation of rainfall is unrelated to plot-level efficiency is rejected with a test statistic of LR = -2(-377.59 – (-356.09)) = 43.00. The regression coefficient on the coefficient of variation of rainfall is positive and significant, confirming it as a reasonable proxy for rainfall variability. A 1% increase in variability increases inefficiency by 0.36%.

The productivity frontier is less robust when the number days of heat stress is interacted with the coefficient of variation of seasonal rainfall (see Table 3). The relative sizes of the input elasticities in the production function sub-model change dramatically compared to the baseline and Model 1. The sign on the coefficient on chemicals becomes negative and heat stress changes from a stress factor to an enhancer of productivity. Another reason for rejecting this specification is that, in this model, the log-likelihood statistic is much lower than that of Model 1.

The analysis was repeated with the standard deviation of rainfall instead of the coefficient of variation, with much the same result. These are not shown. In the equivalent of Model 1, two important coefficients were no longer significantly different from zero, namely chemicals and the drought proxy, but all signs were as expected. The log-likelihood statistic was lower than in Model 1. Interacting heat stress with the standard deviation of rainfall caused fewer problems than it did in Model 2. In the equivalent of Model 2, the coefficient on chemicals was insignificant, although its sign remained positive. The coefficient on heat stress remained positive although the sign on the standard deviation of rainfall became negative and the sign on the interaction term was positive. The log-likelihood statistic was -383.42, an improvement on Model 2, but the interpretation on the coefficient of the interaction term is less straightforward than it had been in Model 2. Therefore, neither coefficient of variation on rainfall nor the standard deviation of rainfall worked particularly well as proxies for periodic drought.

Can rainy days capture periodic droughts or the lack thereof?

Conceptually, rainy days are an enhancer rather than a stressor of crop productivity and the simple count format is easy to interpret. Replacing the coefficient of variation with this count variable produced the results in Table 4. Model 3, the specification with no interaction term between heat stress and rainfall effects, performed almost as well as Model 1. The input elasticities were similar, the inefficiency sub-model's results were as expected and the only difference was that the coefficient on mechanisation was not significantly different in Model 3. Mean efficiency was 56% and the scores increased by 3.43% per year due to the benefits of zero tillage. The magnitude of the heat stress penalty of 1.75% per additional day above 24.9 °C was similar to that in Model 2. That is, each rainy day increases productivity by 1.45%. Because the positive effect of a rainy day was smaller than the negative effect of a heat stress day, the interaction term was expected to carry the same sign as heat stress.

A likelihood ratio test was used to determine if the coefficient on the interaction term in Model 4 should be included. The result was that the coefficient is not zero. The test statistic of Model 3 as a restriction of Model 4 yielded a test statistic of LR = 75.47. This is an anomaly as neither chemicals nor mechanisation produced significant coefficients in Model 4. There were other problems in the inefficiency model too. All four coefficients were significant and the no tillage benefit and heat stress produced the expected signs, but rainy days became a stressor while the combined effect of heat stress and rainy days was positive. While the latter could mean that temperatures above 25 °C are not a problem if there is enough moisture in the soil profile, it is inconceivable that more frequent rainfall on its own would have a negative impact on productivity.

Does an aridity index simplify matters?

The aridity index used in the results in Table 5 was adapted from Dinar et al.[10] In Model 5, the simplest specification, where the aridity index is combined only with the cumulative no tillage benefit, the coefficient on the former is positive and significant. This result is consistent with the result produced by Model 3 in Table 4 because the aridity index rises with heat stress and falls with more frequent precipitation. The differences between the two models are minor. In Model 3 heat stress is defined according to an arbitrary cut-off, while in Model 5 there is no temperature cut-off assumed. With respect to rainfall, Model 5 does not consider the distribution of rainfall, which Model 3 does. In both cases, only three of the four input elasticities are significantly different from zero, but the variables that are not significant are different between the models, including the elasticity on chemicals which is much lower than previously found.

Table 3: Grain productivity model testing the standard deviation of rainfall as drought proxy

	Model 1 – no interaction				Model 2 – interaction			
	Coefficient	**s.e.**	***t*-ratio**	**Significance**	**Coefficient**	**s.e.**	***t*-ratio**	**Significance**
Constant	1.554	0.793	1.96	*	-11.162	3.215	-3.47	***
Seed and fertiliser	0.430	0.066	6.47	***	0.970	0.199	4.88	***
Chemicals	0.084	0.035	2.38	**	-0.213	0.121	-1.76	*
Mechanisation	0.150	0.073	2.04	**	1.546	0.307	5.04	***
Rainfall	0.508	0.046	10.92	***	0.660	0.211	3.13	***
Constant	-35.923	3.849	-9.33	***	-0.601	1.013	-0.59	
No till benefit	-1.595	0.091	-17.45	***	-0.443	0.088	-5.02	***
Heat stress	0.273	0.068	4.00	***	-0.668	0.116	-5.76	***
Coefficient of variation (CV) of rainfall	11.702	1.718	6.81	***	-7.257	0.891	-8.14	***
Heat stress x CV rainfall					0.462	0.042	11.10	***
Sigma-squared	7.288	0.453	16.10	***	2.878	0.290	9.92	***
Gamma	0.998	0.001	1162.85	***	0.872	0.031	28.37	***
Log-likelihood statistic	-356.09				-444.79			
Observations	392				392			

***p ≤ *0.01* , **p ≤ *0.05* , *p ≤ *0.10*

Table 4: Grain productivity model testing the number of rainy days as drought proxy

	Model 3 – no interaction				Model 4 – interaction			
	Coefficient	**s.e.**	***t*-ratio**	**Significance**	**Coefficient**	**s.e.**	***t*-ratio**	**Significance**
Constant	1.534	0.952	1.61		0.738	0.756	0.98	***
Seed and fertiliser	0.462	0.106	4.38	***	0.620	0.072	8.57	***
Chemicals	0.085	0.035	2.43	**	0.053	0.034	1.56	
Mechanisation	0.113	0.076	1.48		0.092	0.061	1.51	
Rainfall	0.511	0.056	9.06	***	0.494	0.037	13.23	***
Constant	-5.311	1.042	-5.10	***	-145.386	27.570	-5.27	***
No till benefit	-1.267	0.126	-10.07	***	-1.874	0.109	-17.17	***
Heat stress	0.536	0.027	19.50	***	4.534	0.762	5.95	***
Rainy days	-0.628	0.054	-11.72	***	6.329	1.267	5.00	***
Heat stress x rainy days					-0.209	0.038	-5.56	***
Sigma-squared	8.931	0.788	11.33	***	13.938	1.058	13.17	***
Gamma	0.999	0.000	2069.28	***	0.999	0.000	2579.67	***
Log-likelihood statistic	-368.47				-330.70			
Observations	392				392			

***p ≤ *0.01* , **p ≤ *0.05* , *p ≤ *0.10*

Table 5: Grain productivity model with aridity index

	Model 5 – no interaction				Model 6 – interaction				Model 7 – interaction			
	Coefficient	**s.e.**	***t*-ratio**	**Significance**	**Coefficient**	**s.e.**	***t*-ratio**	**Significance**	**Coefficient**	**s.e.**	***t*-ratio**	**Significance**
Constant	2.644	0.752	3.51	***	1.786	0.692	2.58	**	1.804	0.870	2.07	**
Seed and fertiliser	0.342	0.078	4.40	***	0.398	0.081	4.91	***	0.456	0.092	4.96	***
Chemicals	0.016	0.041	0.38		0.075	0.039	1.94	*	0.068	0.039	1.76	*
Mechanisation	0.143	0.050	2.87	***	0.180	0.082	2.18	**	0.122	0.084	1.46	
Rainfall	0.543	0.025	22.04	***	0.491	0.044	11.23	***	0.484	0.044	10.95	***
Constant	5.066	0.503	10.07	***	-32.295	1.308	-24.70	***	-5.421	1.052	-5.15	***
No till benefit	-2.761	0.292	-9.45	***	-1.524	0.098	-15.52	***	-1.241	0.143	-8.67	***
Aridity index	9.533	2.312	4.12	***	1.198	1.057	1.13		-1.903	1.023	-1.86	*
Heat stress					0.924	0.033	28.27	***	0.538	0.028	19.06	***
Rainy days									-0.624	0.060	-10.34	***
Sigma-squared	15.869	1.917	8.28	***	10.671	0.791	13.49	***	8.695	0.944	9.21	***
Gamma	1.000	0.000	7309.11	***	0.999	0.000	2804.99	***	0.998	0.001	1355.65	***
Log-likelihood statistic	-432.14				-377.47				-368.98			
Observations	392				392				392			

***$p \leq 0.01$, **$p \leq 0.05$, *$p \leq 0.10$

In Model 6 the heat stress count variable was added. This specification is most like the baseline model in Table 2 because it does not include any rainfall proxy other than total rainfall in the production frontier. Including the heat stress count variable makes the aridity index insignificant, although a likelihood ratio test rejects its exclusion. The test value is LR $= 109.34$. This time all the elasticities are plausible and significant at $p \leq 0.10$, while the coefficient on the aridity index, although still positive, is no longer statistically significant.

The final specification in Model 7 added the number of rainy days recorded during the rainy season, which improved the fit slightly compared to Model 6. The frontier performs well, despite the smaller elasticity on mechanisation and its lack of significance. The results of the inefficiency model are better than all the earlier specifications. More time after the adoption of zero tillage significantly decreases inefficiency, more heat stress increases inefficiency and more rainy days decreases inefficiency. All of this is the same as in Model 3 in Table 4. However, in Model 7, the aridity index provides an interaction effect between rainfall and temperature. The coefficient is significant at $p \leq 0.10$ and negative, which means that inefficiency is negatively correlated with aridity. However, interpreted as an interaction term of rainfall and temperature, the negative sign could mean that the net effect is determined by rainfall and not by heat stress. The low level of significance suggests that the net effect is site specific, which seems eminently reasonable. The mean efficiency and the percentage change in efficiency scores per year of zero tillage, or heat stress or rainy day recorded during the growing season, are stable across specifications.

Conclusions

We used experimental farm data to investigate the impact of climate change on productivity in South Africa, importantly using a new approach for rainfall and temperature stress. The approach followed the Ricardian models used in the literature but extended these in two important respects. Firstly, the data are for a single location, although many different plots were included and the analysis was able to capture changes over time. Secondly, the standard Ricardian specification was estimated using a stochastic frontier with inefficiency effects, which made it possible to allocate the results to the plot level.

The many specifications in which proxies for heat and moisture stress have been used to explain crop plot efficiency and the result provide confidence that this approach is viable and can be extended to encompass other environments and other contexts. The main implications of the results are twofold. Firstly, climate change will have consequences for farm efficiency, and secondly, something can be done about it. This first conclusion is already visible in how plot level efficiencies respond to the temperature and rainfall variation documented at the site over the last decade and a half, regardless of whether this variation is considered normal or a sign of permanent climate change. If either rainfall or temperature is likely to become more variable in future, farm productivity will decrease as a result, and not just become more variable as previously suggested.[18]

The positive results of this research show that, while little can be done to influence rainfall or temperature, at least in the short term, producers can choose their production system. In the Langgewens experiment, the first 14 years of zero tillage production practices resulted in an almost 3.5% increase in productivity every year, across good and bad seasons. This effect is unlikely to be linear, and so the marginal efficiency benefits will probably decline as soil benefits mature, and finding the shape and length of the lag structure involved is a topic for further research. Certainly, it is not known at this stage how long the benefits take to mature and there is a difference in the rate at which the benefits accumulate across different rotation systems. Finally, from a modelling point of view, efficiency can replace crop yield or its derivatives in climate models, and the Dinar aridity index is worth pursuing; capturing rainfall variability with a coefficient of variation does not work quite as well as a simple count variable. All of these preliminary conclusions can be usefully re-examined in other contexts. What is without doubt is that careful collection of climate data along with production variables is essential.

Acknowledgements

We acknowledge the input of the two anonymous reviewers.

Competing interests

We have no competing interests to declare.

Authors' contributions

B.C.: Conceptualisation, data cleaning, empirical analysis, first draft. J.P.: Conceptualisation, final draft. J.S.: Data collection.

Data availability

Data are available on request from Johann Strauss (JohannSt@elsenburg.com) and subject to permission from the Western Cape Department of Agriculture.

References

1. Gbetibouo GA, Hassan RM. Measuring the economic impact of climate change on major South African field crops: A Ricardian approach. Glob Planet Change. 2005;47(2–4):143–152. https://doi.org/10.1016/j.gloplacha.2004.10.009
2. Water Research Commission, South African Weather Service. A climate change reference atlas based on CMIP5-CORDEX downscaling [document on the Internet]. c2017 [cited 2019 Aug 12]. Available from: http://www.weathersa.co.za/Documents/Climate/SAWS_CC_REFERENCE_ATLAS_PAGES.pdf
3. Mendelsohn R, Nordhaus WD, Shaw D. The impact of global warming on agriculture: A Ricardian analysis. Am Econ Rev. 1994:753–771.
4. Sanghi A, Mendelsohn R, Dinar A. The climate sensitivity of Indian agriculture. In: Dinar A, Mendelsohn R, Evenson R, Parikh J, Sanghi A, Kumar K, et al. (eds). Measuring the impact of climate change on Indian agriculture. World Bank Technical Paper number 402. Washington DC: World Bank; 1998. p. 69–139.
5. Kurukulasuriya P, Mendelsohn R, Hassan R, Benhin J, Deressa T, Diop M, et al. Will African agriculture survive climate change? World Bank Econ Rev. 2006;20(3):367–388. https://doi.org/10.1093/wber/lhl004
6. Massetti E, Mendelsohn R. Estimating Ricardian models with panel data. Clim Chang Econ. 2011;2(04):301–319. https://doi.org/10.1142/s2010007811000322
7. Mwaura FM, Okoboi G. Climate variability and crop production in Uganda. J Sustain Dev. 2014;7(2):159. https://doi.org/10.5539/jsd.v7n2p159
8. Cabas J, Weersink A, Olale E. Crop yield response to economic, site and climatic variables. Clim Change. 2010;101(3–4):599–616. https://doi.org/10.1007/s10584-009-9754-4
9. Boubacar I. The effects of drought on crop yields and yield variability: An economic assessment. Int J Econ Finance. 2012;4(12):51–60. https://doi.org/10.5539/ijef.v4n12p51
10. Dinar A, Karagiannis G, Tzouvelekas V. Evaluating the impact of agricultural extension on farms' performance in Crete: a nonneutral stochastic frontier approach. Agricultural Economics. 2007 Mar;36(2):135-46. https://doi.org/10.1111/j.1574-0862.2007.00193.x
11. Roux PW. South Africa devises scheme to evaluate drought intensity. Drought Network News. 1991;3(3):18–23.
12. MacLaren C, Storkey J, Strauss J, Swanepoel P, Dehnen-Schmutz K. Livestock in diverse cropping systems improve weed management and sustain yields whilst reducing inputs. J Appl Ecol. 2019;56(1):144–156. https://doi.org/10.1111/1365-2664.13239
13. Battese GE, Coelli TJ. A model for technical inefficiency effects in a stochastic frontier production function for panel data. Empir Econ. 1995;20(2):325–332. https://doi.org/10.1007/bf01205442
14. Piesse J, Conradie B, Thirtle C, Vink N. Efficiency in wine grape production: Comparing long-established and newly developed regions of South Africa. Agric Econ. 2018;49(2):203–212. https://doi.org/10.1111/agec.12409
15. South African Department of Agriculture Fisheries and Forestry (DAFF). Abstract of agricultural statistics. Pretoria: DAFF; 2018.
16. Conradie B, Piesse J, Thirtle C. District-level total factor productivity in agriculture: Western Cape Province, South Africa, 1952–2002. Agric Econ. 2009;40(3):265–280. https://doi.org/10.1111/j.1574-0862.2009.00381.x
17. Battese GE, Corra GS. Estimation of a production frontier model: with application to the pastoral zone of Eastern Australia. Aust J Agric Econ. 1977;21(3):169–179. https://doi.org/10.1111/j.1467-8489.1977.tb00204.x
18. Speranza CI, Kiteme B, Wiesmann U. Droughts and famines: The underlying factors and the causal links among agro-pastoral households in semi-arid Makueni district, Kenya. Glob Environ Change. 2008;18(1):220–233. https://doi.org/10.1016/j.gloenvcha.2007.05.001

AUTHORS:
Abel L. Mtembeji[1,2]
Dharam R. Singh[1]

AFFILIATIONS:
[1]Division of Agricultural Economics, ICAR-Indian Agricultural Research Institute, New Delhi, India
[2]Dar es Salaam City Council, Dar es Salaam, Tanzania

CORRESPONDENCE TO:
Abel Mtembeji

EMAIL:
abell82@yahoo.com

DATES:
Received: 30 Apr. 2021
Revised: 23 June 2021
Accepted: 28 June 2021
Published: 29 Sep. 2021

HOW TO CITE:
Mtembeji AL, Singh DR. Dynamics of rice production among the food crops of Tanzania. S Afr J Sci. 2021;117(9/10), Art. #11020. https://doi.org/10.17159/sajs.2021/11020

ARTICLE INCLUDES:
☒ Peer review
☐ Supplementary material

DATA AVAILABILITY:
☐ Open data set
☐ All data included
☒ On request from author(s)
☐ Not available
☐ Not applicable

EDITORS:
Terresa Coutinho
Salmina Mokgehle

KEYWORDS:
rice production, compound annual growth rate, instability, Markov chain, Tanzania

FUNDING:
None

Dynamics of rice production among the food crops of Tanzania

Rice is an important crop in Tanzania which contributes significantly to the farmers, consumers, and the government. Recognising this importance, the government has made initiatives to attain rice self-sufficiency. These initiatives are crucial in contributing to regional self-sufficiency, enabling rice market leadership, and injecting productivity through significant improvements in the quality, quantity, and value of rice produced in Tanzania. We investigated the dynamics of rice area, production, and productivity and identified shifts in the land-use patterns in Tanzania. To analyse secondary data collected over a 33-year period from 1986/1987 to 2018/2019, we used compound annual growth rates, Cuddy-Della Valle Index and a first-order Markov chain approach. We found that the growth in the areas under rice cultivation, production and productivity were inconsistent as evidenced by the presence of instabilities. Rice remains the third most stable crop in the country in terms of area under production retention; however, this might decline in the next 2 years. Policies in future must enable strategies to increase productivity as well as promote high-yielding varieties, efficient input usage, and irrigation infrastructure development.

Significance:

- We investigated the spatial and temporal trends in rice area, production, and productivity as well as identified shifts in the crop land-use patterns after a series of government interventions. To achieve sustainability, it is essential to revisit agricultural crop growth strategies regularly at macro- and micro-levels.

Introduction

Agriculture is an important economic sector in Tanzania, contributing about 30% to the country's total GDP[1] of USD47.43 billion. It is the primary source of food and jobs (65.5%) in the country, besides supplying raw materials for industries (65%) and contributing to the foreign exchange earnings (30%). Agricultural production is a critical component of Tanzania's Poverty Reduction Strategy, as it contributes to the country's long-term economic development and food security.[2] Nevertheless, it is affected by several factors, including crop management, climate, socio-economic factors, and the policies of the government. Production variability also results from variations in region, yield, and/or the interaction of area and yield. As a result, with time, it is crucial to understand the nature and patterns of the area, production, and productivity to identify policy interventions that can help close the yield gap and improve sustainability.

While the agriculture sector of Tanzania is growing at 5.2%, the crop sector grows at 5.8%. Cereals, pulses, roots and tubers, fruits and vegetables, and spices are among the major crops grown. Cereals account for 52% of the total dietary energy supply, with maize and rice accounting for 59% and 21%, respectively.[3] In Africa, rice consumption has been growing at a faster rate than any other staple food, at around 5.5% per year, due to urbanisation and related changes in eating habits, as well as population growth.[4] In Tanzania, it is the second most common cereal after maize, and it contributes to the country's food and household nutrition security. It is grown in 64 districts, affecting the livelihoods of over two million people.[3,5] Like maize, it is a strategic crop for Tanzanian government agricultural investments, as it contributes to the value of agricultural production, national food security and export revenues.[5]

Tanzania has developed a range of agricultural development initiatives over the last 30 years, including the National Agriculture Policy[6], which serves as a guideline for the overall development of the agriculture sector, with a focus on crop production optimisation for food security and economic development. Programmes, initiatives, policies, strategies, and projects carry out the policy implementation. In the medium and long term, the Agricultural Sector Development Strategy of 2003–2013 provided an encouraging and cooperative environment for improving agricultural productivity and profitability as the foundation for improved farm incomes and rural poverty reduction. Meanwhile, the Agricultural Sector Development Programme I of 2005–2015 served as the agriculture sector's overall mechanism for overseeing the sector's structural, expenditure, and investment development.[7] The 2015–2025 Agricultural Sector Development Strategy II set a target of contributing to Tanzania's national economic growth and poverty reduction (Vision 2025/LTPP) by promoting inclusive and sustainable agricultural growth, reducing rural poverty by 2025/2026, and improving food and nutrition security (by reducing the per cent of rural households below the food poverty line by 2025/2026). Furthermore, the Agricultural Sector Development Programme II of 2017–2027 is a tool for strategy execution, to enhance smallholder farmers' productivity of target commodities within sustainable production systems and to forge sustainable market linkages for competitive surplus commercialisation and value chain development.[8]

Rice is the world's most commonly grown cereal grain, and it is a staple food for more than 60% of the world's population.[9] China is the world's largest rice producer, accounting for nearly 30% of global output; other major producers include India, Indonesia, Bangladesh, Vietnam, Thailand, Burma, Philippines, and Brazil. Between 1961 and 2010, global rice production more than tripled, with an annual compound growth rate of 2.24% (2.21% in rice-

producing Asia). Much of the rise in rice production is attributed to higher yields, which rose at an annual average rate of 1.74%, compared with an annual average growth rate of 0.49% for the area harvested. Paddy yields rose at an annual average rate of 51.1 kg/ha per year in absolute terms.[10] The rice industry is contributing significantly to farmers, consumers, and the government of Tanzania. Several government-led initiatives were implemented for the development of the sector. For example, the implementation of the 10-year first phase of the National Rice Development Strategy of 2009 aimed to double rice production by 2018[11], and the National Rice Development Strategy II of 2019 was launched to support the government's rice development efforts over the next 12 years. The goal is to maintain rice self-sufficiency, contribute to regional self-sufficiency, enable rice market leadership, and inject productivity through significant improvements in the quality, quantity, and value of rice produced in Tanzania.[5]

Compound annual growth rate (CAGR) has become a common tool for analysing the growth in crop production. For example, CAGR was used to measure the change and growth rate in the area, production, and yield of wheat in Ethiopia from 1991/1992 to 2012/2013.[12] In addition, an examination of the growth and trend in plantain area, production quantity, and yield in Nigeria between 1961 and 2017[13], as well as an analysis of the contributions of yield and area to the production of cassava in Nigeria[14], used CAGR. Findings suggest that improved productivity is important, because expanding area under a specific crop might not be feasible without reducing the area share of other crops. Thus, it is crucial to deploy and implement suitable technologies.[14] When CAGR is paired with instability indices, it is possible to obtain a deeper understanding of the degree of continuity in crop growth and to classify the winning variable if growth comes first then a decrease in volatility.[15-17] Assessing improvements in cropping patterns in various regions is also important for obtaining a deeper understanding of the agricultural growth and development process. A first-order Markov chain analysis provides a base for study into the complexities of cropping trends and patterns by examining the area retention (stability of the area under various crops) and crop shifting from one set to another.[18-20]

Despite many studies on the area, production, and productivity for major food crops undertaken for Tanzania in the past, there have been very few nationwide studies on rice in recent years. Mkonda and He[21] used the Mann–Kendall test to assess the development pattern of major food crops (five crops) and their efficacy on food security in Tanzania from 1980 to 2015. During the study period, they discovered that production had a positive trend while yield had a negative trend. The supply response, price and non-price factors determining output, and how receptive farmers were to these factors were examined through a panel survey in 2008/2009.[2] Some core determinants of production were established and demonstrated that farmers are price responsive. In this study, we investigated the dynamics of rice crop area, production, and productivity in Tanzania over a 33-year period (1986/1987–2018/2019) in three phases, and we identified shifts in the land-use patterns of rice in relation to other major food crops (maize, sorghum, millets, wheat, pulses and others) as indicators of sustainability.

Methodology

To examine trends in rice area, production and productivity for a 33-year period (1986/1987 to 2018/2019), we used secondary data from the Ministry of Agriculture of Tanzania, the National Bureau of Statistics (Tanzania), and FAOSTAT. The study period was divided into three sub-periods (1986/1987–1998/1999, 1999/2000–2008/2009, and 2009/2010–2018/2019), totalling 33 years (1986/1987–2018/2019). The sub-period wise analysis was done to assess the effect of the rice policy intervention (National Rice Development Strategy I) implemented between 2009 and 2018. This period was sufficient to gain insight into the production performance and establish plans for potential initiatives. The following methods were used for the analysis.

Trend growth

A variable's growth reflects how well it has performed in the past. The growth analysis determines the trend of a given variable over time. The exponential growth function in the form of a CAGR was used to estimate the growth in the area, production, and productivity of the rice (paddy) crop in Tanzania as it was applied by Bezabeh et al.[12]; Adeoye et al.[13]; Ikuemonisan et al.[14]; Ayalew[16]; Bisht and Kumar[17]. CAGRs were calculated by exponentially regressing the time series data on area, production and productivity of rice against time using the following formula:

$$Y_t = ab^t e^u \qquad \text{Equation 1}$$

where Y_t is the dependent variable for which growth rate was estimated (area, production and productivity); a is the intercept; $b=(1+r)$; r is the annual growth rate; t is the year (time period) which takes values 1,2, 3, …n; and u is the error term for the year t.

The equation was transformed into log-linear form and estimated using the ordinary least squares method. The CAGR (per cent) was then calculated from the relationship CAGR (%) = (antilog of ln b-1)*100. The estimation was done for all three phases (periods) – 1986/1987–1998/1999, 1999/2000–2008/2009, 2009/2010–2018/2019 – and the overall period of study, 1986/1987–2018/2019.

Instability

The Cuddy-Della Valle Index, which is a measure of variability in time series data, was used to identify instabilities in rice area, production and productivity.[22] It is an improvement over the simple coefficient of variation, which is prone to overestimation. For a series with a time trend, the Cuddy-Della Valle Indexes for three periods (1986/1987–1998/1999, 1999/2000–2008/2009, 2009/2010–2018/2019) and for the overall period of study (1986/1987–2018/2019) were calculated as follows:

$$\text{Cuddy-Della Valle Index (\%)} = CV \times \sqrt{(1-\bar{R}^2)} \qquad \text{Equation 2}$$

where CV is the coefficient of variation in per cent and $\bar{R}^2$ is the adjusted coefficient of determination from a time trend regression.

Markov chain analysis

Secondary data for the major food crops were obtained from the Ministry of Agriculture (Department of Food Security), United Republic of Tanzania, to investigate the dynamics of the shift in cropping patterns for the 10-year period (2009/2010–2018/2019). Analysis using these data aided in understanding the performance of the rice crop in relation to other food crops. The major food crops investigated were maize, sorghum, millets, rice, wheat, pulses and others (cassava, banana, Irish potatoes and sweet potatoes). External non-stationary factors such as rainfall, temperature, and agricultural input and output prices influence farmer crop selections. The majority of the rice produced in the country is under rainfed conditions.[4] There is a stiff competition with other food crops in terms of area under production because all the selected food crops grow during the long rainy season (agricultural season). As a result, all the major food crops compete for land and inputs based on these factors' variations. We used a first-order Markov chain approach to evaluate the direction of shift in cropping pattern using LPSOLVE IDE software and Microsoft Excel.

The transitional probability matrix 'P', whose diagonal elements represent retention probability and off-diagonal elements represent switching-over probability, was examined using a first-order Markov chain analysis. The diagonal elements of the analysis show how much land a crop had maintained from previous years. The higher the value in the diagonal element (which tends to 1), the more stable the crop is, whereas lower values (which tend to 0) suggest instability. The crop's column elements represent the area gained from other crops, while the crop's row elements indicate the area lost to other crops.[18-20] The average area shifted to a particular crop is a random variable that depends only on the previous crop's area, algebraically expressed as:

$$E_{jt} = \sum_{i=1}^{n} (Eit\text{-}1)\, Pij + e_j \qquad \text{Equation 3}$$

where E_{jt} is the area of the crop shifted towards the particular jth crop in the year t; E_{it}-1 is the area lost by the *i*th crop during the year *t*-1; P_{ij} is the probability that the area lost will shift from *i*th crop to the *j*th crop; E_{jt} is the error term which is statistically independent of the E_{it}-1; and *n* is the number of crops.

The transitional probabilities *pij*, which can take a (c × r) matrix form, have the following properties:

1. $0 < P_{ij} < 1$

2. $\sum P_{ij} = 1$, for all *i*

The following formulae were used to project the areas for the seven crops for the next 2 years (2019/2020–2020/2021) based on the results of the Markov chain analysis:

$$B_t = B_0 x T \quad \text{Equation 4}$$

$$B_{(t+1)} = B_{(t+i-1)} x T \quad \text{Equation 5}$$

where B_0 is the area under the crop in the base year; $B_{(t+1)}$ is the area under the crop in the next year (prediction); and *T* is the transitional probability matrix.

Results and discussion

Tanzania has a long tradition of rice production, and its rice agroecosystems are divided into three categories: irrigated lowland, rainfed lowland, and rainfed upland, with the rainfed lowland areas producing the majority of the rice.[4,23] The potential area for irrigation is 29.4 million hectares, but the current paddy area under irrigation is 475 052 ha, which increased from a figure of 289 245 ha in the year 2007/2008.[5,11] Domestically, fertiliser availability has also improved, from 302 000 tonnes in 2009/2010 to 418 355 tonnes in 2018/2019. Farmers in the country cultivate hundreds of local rice varieties, as well as a few improved varieties such as SUPA, NERICA, TXD 88, TXD 306 (SARO 5).[4,5]

Rice imports are projected to decrease because of the Tanzanian government's ban on rice imports in 2018 and a rise in domestic production to the point that the country can maintain rice self-sufficiency. Long-grain milled rice is currently imported from Pakistan, Thailand, and India. According to the Ministry of Agriculture, rice imports have fluctuated in recent years, from 8 tonnes in 2013, 9069 tonnes in 2014, 28 888 tonnes in 2015, 959 043 tonnes in 2016, to 942.5 tonnes in 2017.

Tanzania has realised a rise in rice production and area under cultivation over the 33-year study period (1986/1987 to 2018/2019). Figure 1 shows that the highest level of production (2 229 071 tonnes) was in 2015/2016 and the lowest was in 1991/1992 (256 000 tonnes), with the lowest area under production (200 090 ha) in 1991/1992 and the highest in 2009/2010 (1 141 065 ha).

Rice productivity ranged considerably, with 1 tonne per hectare being the lowest and 2 tonnes per hectare being the highest. This productivity is far below that of the top ten producing countries in the world, with the top three producers, China, Japan, and Brazil, producing yields of 7.1, 6.8, and 6.1 tonnes per ha, respectively, as shown in Figure 2.

Mbeya, Morogoro, Mwanza, Tabora, Shinyanga, and Rukwa regions are the country's major rice producers and are thus the 'big six' of rice production. In 2018/2019, the six regions produced 62% of the country's total rice production (their individual contributions are reflected in Figure 3), accounting for 61% of the 1 052 547 ha under cultivation. Rice productivity was found to be higher than the national average (1.8 tonnes/ha) in two of the big six regions (Mbeya and Morogoro). Within the same year, Mbeya and Morogoro produced 3.0 tonnes/ha and 2.5 tonnes/ha, respectively. If sufficient efforts, including technological advancements, are maintained in the big six regions, it is possible to surpass the national average in productivity.

Compound annual growth rate

To understand the growth patterns of the area, production, and productivity of paddies, the CAGR was calculated for each variable in the three separate periods as well as overall for the 33-year period, as shown in Table 1. Over a 33-year period, the area under production grew at a significant rate of 7.48% in the first period and 8.44% in the second, and overall, the area under production grew at a rate of 6.22% annually. The annual growth of the area under production has shown a swing periodically for the past 33 years, increasing from 7.48% to 8.44%, and then growing insignificantly in the third period.

The CAGRs of production were found to be 4.97%, 6.41% and 4.8% in the first, second and third periods, respectively, indicating that the rate of production growth has recently slowed. Over the entire study period, production grew significantly at a rate of 5.81% per year. The productivity growth was significant for the first and third periods with CAGRs of -2.33% and 4.73%, respectively. The productivity grew positively and significantly at 4.73%, implying that Tanzania has only experienced significant productivity growth in the last 10 years (2009/2010–2018/2019). Over the entire study period (1986/1987–2018/2019), there was a substantial CAGR of 6.22% for the area and 5.81% for production, as also observed for trends of rice area and production in Andhra Pradesh of India[26], suggesting the potential to target initiatives aimed at increasing productivity growth.

Instabilities of rice production

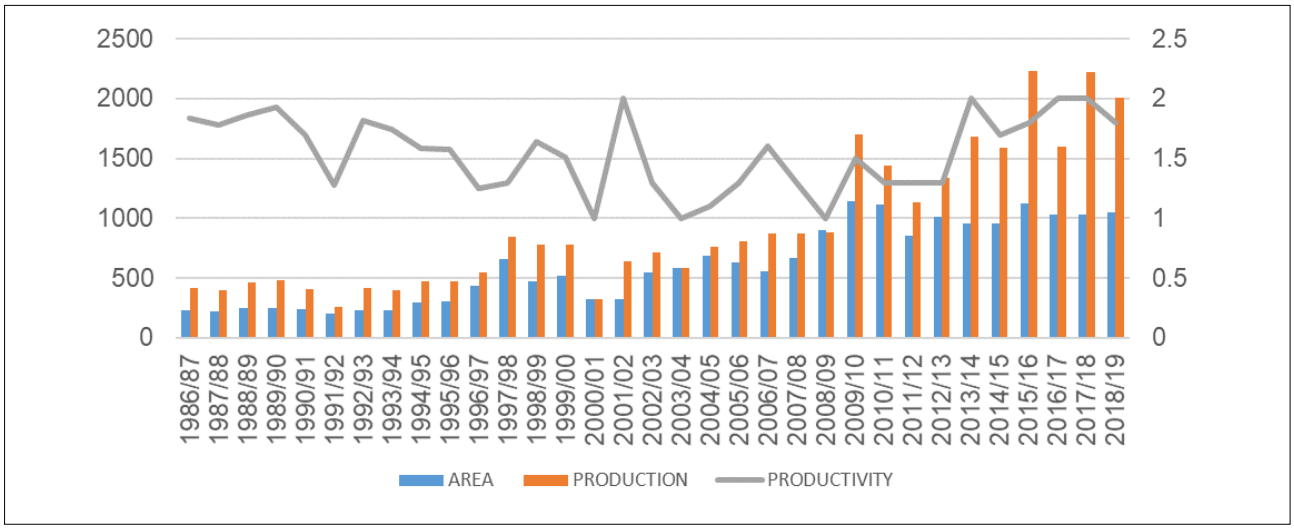
The Cuddy-Della Valle Index for the area, production, and productivity instability was calculated and the results are shown in Table 2 to better understand the consistency of growth performance. Throughout the study period, area instability decreased from 29.73%, 19.14% and 9.7% in the first, second, and third periods, respectively. The first two stages showed medium instability, while the third period showed low instability. This signifies that the variations of area under rice production decreased over time. The area overall had a medium instability index of 20.28% over the 33-year period (1986/1987–2018/2019).

Production instability has decreased significantly over time – 25.27%, 18.55% and 17.07% for the first, second, and third periods, respectively – but has remained at a medium level of instability. Production gained stability in the third period in comparison to the first and second periods, which indicates a decrease in production variations. Production instability was moderate over the entire 33-year period, at 29.61%. Instability in production intensifies price swings, making disadvantaged farmers more vulnerable to market forces. Productivity instability was low in the first period (11.34) and third period (12.39), whereas it was moderate in the second period (25.22). In general, a moderate instability was observed for productivity over the entire study period considered, at 20.69. Productivity instability was slightly higher than area instability, indicating that productivity variation was a major source of variation in rice production in Tanzania, as also found by Mkonda and He[21] for the period 1980–2015.

Dynamics of the shift in cropping patterns of major food crops

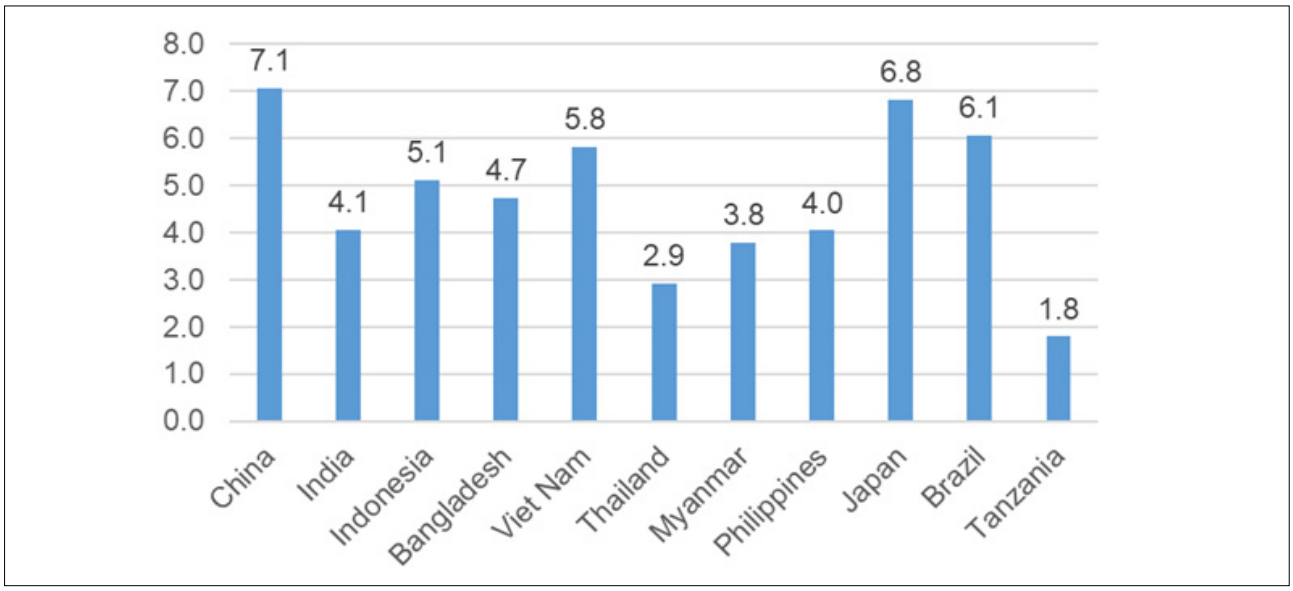
Table 3 shows the transition probability matrix for major crops during the study period that indicates the proportion changes of area (ha) under production from year to year. The wheat crop was found to be the most stable in terms of area retention for the period 2009/2010–2018/2019 by retaining 56% of its previous area (ha), followed by other food crops (cassava, banana, Irish potatoes and sweet potatoes) which retained 49%. The rice crop was the third most stable crop with 39.2%, followed by maize with 39.1%. Millets were the only crop group that did not retain the previous area under cultivation, which signifies that millet farmers are shifting the area and looking to cultivate other crops.

Rice displayed properties of a dynamic crop in terms of area under cultivation as the findings show that it gained 17% area from maize and lost 15% and 45.5% to pulses and other crops, respectively. The stability of the area under production for rice is still low, which is the result of having multiple food crops in the country which encourage farmers to switch to different crops to attain better outputs.



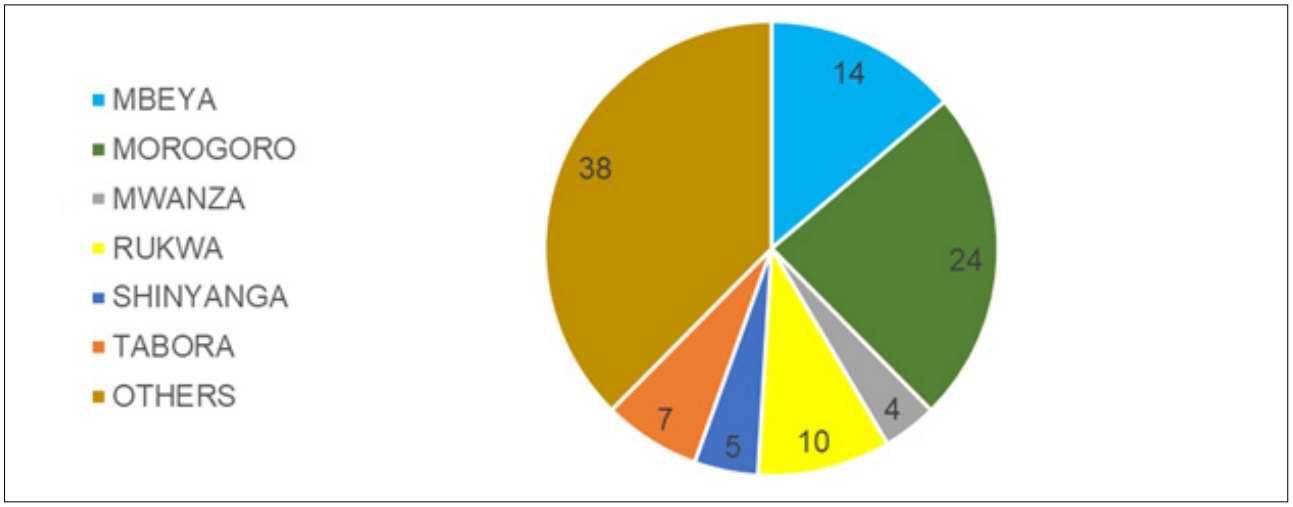
Data source: Tanzanian Ministry of Agriculture[24]

Figure 1: Production of rice in Tanzania.



Data source: FAOSTAT[25]

Figure 2: Comparison of yield/productivity (tonnes/ha) of major rice-producing countries and Tanzania for the year 2019.



Data source: Tanzanian Ministry of Agriculture[24]

Figure 3: Region-wise share (%) of rice production in Tanzania for 2018/2019.

Table 1: Compound annual growth rate (CAGR) for rice production in Tanzania

Period	Area		Production		Productivity	
	Coefficients	CAGR (%)	Coefficients	CAGR (%)	Coefficients	CAGR (%)
1986/1987–1998/1999	0.0721* (0.2300)	7.48	0.0485** (0.2471)	4.97	-0.0236** (0.1230)	-2.33
1999/2000–2008/2009	0.081* (0.0237)	8.44	0.0621*** (0.0278)	6.41	-0.0177 (0.0263)	-1.75
2009/2010–2018/2019	-0.0002 (0.0104)	-0.02	0.04686*** (0.0193)	4.8	0.0462* (0.0141)	4.73
1986/1987–2018/2019	0.0584* (0.2119)	6.22	0.0565* (0.2522)	5.81	-0.0013 (0.2175)	-0.13

Data source: Tanzanian Ministry of Agriculture[24]

**, **, ***Indicate 1%, 5%, 10% levels of significance, respectively, and figures in parentheses are standard errors.*

Table 2: Instability index (Cuddy-Della Valle Index) of rice production in Tanzania

	1986/1987–1998/1999	1999/2000–2008/2009	2009/2010–2018/2019	1986/1987–2018/2019
Area	29.73	19.14	9.7	20.28
Production	25.27	18.55	17.07	29.61
Productivity	11.34	25.22	12.39	20.69

Data source: Tanzanian Ministry of Agriculture[24]

Table 3: Transitional probability matrix for the shift in cropping pattern of major food crops of Tanzania during 2009/2010–2018/2019

	Maize	Sorghum	Millets	Rice	Wheat	Pulses	Others	Sum
Maize	0.3905	0.1682	0.0562	0.1607	0.0013	0.1331	0.0900	1
Sorghum	0.5340	0.0212	0.0837	0.0000	0.0185	0.3331	0.0096	1
Millets	0.0000	0.2006	0.0000	0.0000	0.0000	0.0000	0.7994	1
Rice	0.0000	0.0000	0.0000	0.3917	0.0000	0.1538	0.4545	1
Wheat	0.0000	0.0000	0.4402	0.0000	0.5598	0.0000	0.0000	1
Pulses	0.9750	0.0000	0.0000	0.0000	0.0000	0.0250	0.0000	1
Others	0.1608	0.0000	0.0000	0.0000	0.0102	0.3354	0.4935	1

Data source: Tanzanian Ministry of Agriculture[24]

Table 4: Projected area of major food crops of Tanzania (000' ha)

Period	Maize	Sorghum	Millets	Rice	Wheat	Pulses	Others
Base	3428.63	646.87	269.97	1052.55	42.18	1507.10	1929.70
2019/2020	3463.87	644.65	265.36	963.22	59.96	1518.60	1961.33
2020/2021	3492.75	649.61	274.98	933.89	70.24	1519.71	1935.81

Data source: Tanzanian Ministry of Agriculture[24]

The area projections of the seven major food crops for the two periods based on the transitional probability matrix is shown in Table 4. The projected area under rice shows a decreasing trend, which is the same as for the others group of crops. The area under rice production is projected to decrease from 1 052 550 ha of the base in the period 2018/2019 to 963 220 ha for 2019/2020 and 933 890 ha for the period 2020/2021. Rice being a dynamic crop for area retention, the decrease is highly associated with cropping diversification, i.e. farmers have many choices for switching food crops. As the future likelihood depends on the current state, the decline in rice area under production could be a result of variations in non-stationary factors of production and other associated factors including land tenure management, rainfall (due to the risk posed by climate change), shortage of inputs, pests, diseases, access to agriculture extension, access to agricultural credits and marketing problems.[27,28] In addition, the massive migration of labour from agriculture has an impact on labour-intensive crops like rice.[29] Maize, sorghum, millets, wheat, and pulses are projected to grow slightly in area. Implementation of interventions aimed at enhancing productivity are important to increase rice production.

Conclusion

Rice in Tanzania experienced a growth in the area under production at a significant rate of 7.48% in the first period and 8.44% in the second period, while in general, the area under production grew at a rate of 6.22% annually over the 33-year period. The CAGRs of production were 4.97%, 6.41% and 4.8% in the first, second and third periods, respectively.

Over the 33-year period, production grew at a rate of 5.81% per annum. The CAGR of productivity was significant at -2.33% in the first period and 4.73% in the third period. Despite the observed growth in area under production, production and productivity, the growth is inconsistent, as indicated by the presence of instabilities.

Throughout the study period, area instability decreased from 29.73, 19.4, and 9.7 in the first, second, and third periods, respectively. Production instability varied over time, from 15.27, 18.55, and 17.07 for the first, second, and third periods, respectively. Productivity instability was low in the first period (11.34) and third period (12.39), whereas it was medium in the second period (25.22). In general, instabilities for the area, production and productivity of rice were 19.06, 27.13 and 21.11, respectively, in a 33-year duration. Productivity instability was slightly higher than area instability, indicating that productivity variation was a major source of variation in rice production in Tanzania.

Rice is the third most stable crop in the country in terms of area under production retention over the previous 10 years. It retained 39.2% from its previous year's area and lost 15.38% to pulses and 45.45% to other food crops, while it gained 16% from maize. The area under rice production is projected to decrease from 1 073 486 ha in the period 2019/2020 to 1 049 037 ha for the period 2020/2021 and 1 034 204 ha for the period 2021/2022, as a result of variations in non-stationary factors of productions like rainfall (drought), shortage of inputs, pests, diseases and marketing problems. Overall, Tanzanian rice production has been growing over the entire period of analysis, which is mostly a result of the expansion in the area under production. Because of the productivity increase over the last 10 years, the future looks bright. The productivity growth is a result of various initiatives, including the National Rice Development Strategy, that were put in place by the government and other stakeholders.

With the rice area under production projected to fall, the Tanzanian authorities must reconsider their policy measures to align with the growth in rice consumption. The policies, programmes and initiatives must focus on increasing productivity and reducing post-harvest losses so that the available quantity for consumption will increase. The development and promotion of the use of high-yielding varieties, optimum input use, and development of rice irrigation infrastructures might increase the productivity of rice and sustainable food self-sufficiency.

Acknowledgements

This work is based on the PhD research of A.L.M. in the division of Agricultural Economics of ICAR-Indian Agricultural Research Institute.

Competing interests

We have no competing interests to declare.

Authors' contributions

A.L.M.: Conceptualisation, methodology, data collection, data analysis, writing – initial draft. D.R.S.: Conceptualisation, methodology, validation, writing revisions, student supervision.

References

1. Bank of Tanzania (BOT). Annual report 2018/19 [webpage on the Internet]. c2019 [cited 2021 Feb 05]. Available from: https://www.bot.go.tz/Publications/Filter/34
2. Leyaro V, Boulay B, Morrissey O. Food crop production in Tanzania: Evidence from the 2008/09 National Panel Survey Research report to IGC Tanzania number F-40110-TZA-1. London: International Growth Center; 2014.
3. National Bureau of Statistics (NBS). Tanzania in figures, a 2018 government document. Dodoma: NBS; 2018. Available from: https://www.nbs.go.tz/index.php/en/tanzania-in-figures/422-tanzania-in-figures-2018
4. Sekiya N, Oizumi N, Kessy TT, Fimbo KMJ, Tomitaka M. Importance of market-oriented research for rice production in Tanzania. A review. Agron Sustain Dev. 2020;40(1), Art. #7. https://doi.org/10.1007/s13593-020-0611-1
5. United Republic of Tanzania (URT). National Rice Development Strategy Phase ii (NRDS II) 2019-2030. Dodoma: URT; 2019.
6. United Republic of Tanzania (URT). Agricultural and livestock policy 1997. Dar es Salaam: URT; 1997.
7. United Republic of Tanzania (URT). Agricultural sector development programme (ASDP) Support through Basket Fund. Dar es Salaam: URT; 2006.
8. United Republic of Tanzania (URT). Agricultural sector development programme Phase two (ASDP II). Dar es Salaam: URT; 2016.
9. Saito K, Dieng I, Toure AA, Somado EA, Wopereis MCS. Rice yield growth analysis for 24 African countries over 1960–2012. Glob Food Sec. 2015;5:62–69. http://dx.doi.org/10.1016/j.gfs.2014.10.006.
10. CGIAR Research Program on Rice. Rice productivity [webpage on the Internet]. c2021 [cited 2021 Mar 18]. Available from: http://ricepedia.org/rice-as-a-crop/rice-productivity
11. United Republic of Tanzania (URT). National rice development strategy (NRDS). Dar es Salaam: URT; 2009.
12. Bezabeh E, Haregewoin T, Haile GD, Daniel F, Belay B. Change and growth rate analysis in area, yield and production of wheat in Ethiopia. Int J Dev Res. 2014;4(10):1994–1995.
13. Adeoye IB, Olanrewaju O, Adegbite O, Balogun OL. Growth and trend analysis of area, production and yield of major fruits in Nigeria. In: Proceedings of 37th annual conference of the Horticultural Society of Nigeria (HORTSON) "OWO MADE 2019"; 2019 November 18–22; Lafia, Nigeria. Horticultural Society of Nigeria; 2019. p. iii–vii.
14. Ikuemonisan ES, Mafimisebi ET, Ajibefun K. Cassava production in Nigeria: Trends, instability and decomposition analysis (1970–2018). Heliyon. 2020;6(10), e05089. https://doi.org/10.1016/j.heliyon.2020.e05089
15. Nisha Nain M, Aneja DR. Growth analysis of area, production and productivity of wheat crop in Haryana and India. J Exp Biol Agric Sci. 2019;7(3):266–272.
16. Ayalew B. Trends, growth and instability of finger millet production in Ethiopia. Res J Agric Environ Manag. 2015;4(2):78–81.
17. Bisht A, Kumar A. Growth and instability analysis of pulses production in India. Int J Agric Sci. 2018;10(14):6722–6724.
18. Ardeshna NJ, Shiynai RL. Dynamics of cropping pattern in Gujarat state: A Markov chain approach. Asian Acad Res J Soc Sci Humanit. 2013;1(9):56–66.
19. Devi DAR, Reddy RU, Madavi B, Ravi P, Sadvi P. Dynamics of cropping pattern in Karimnagar District of Telangana – A Markov chain approach. Asian J Agric Ext Econ Sociol. 2019;37(4):1–5. https://doi.org/10.9734/AJAEES/2019/v37i430281
20. Jalikatti V, Poddar RS. An economic analysis of changing cropping pattern in Almatti Command Area of Karnataka, India. Int J Curr Microbiol Appl Sci. 2019;8(10):1052–1059. https://doi.org/10.20546/ijcmas.2019.810.123
21. Mkonda MY, He X. Agricultural history nexus food security and policy framework in Tanzania. Agric Food Sec. 2018;7:75. https://doi.org/10.1186/s40066-018-0228-7
22. Bezabeh E. Trends, growth and instability of teff production in Ethiopia. Int J Recent Res Life Sci. 2016;3(4):10–13.
23. Mkanthama J, Makombe G, Kihoro J, Ateka EM, Kanjere M. Technical efficiency of rainfed and irrigated rice production in Tanzania. Irrig Drain. 2018;67(2):233–241. https://doi.org/10.1002/ird.2185
24. Tanzanian Ministry of Agriculture. Tanzania publications [webpage on the Internet]. c2021 [cited 2021 Mar 15]. Available from: https://www.kilimo.go.tz/index.php/en/resources/category/statistics
25. FAOSTAT. Food and agriculture data [webpage on the Internet]. c2021 [cited 2021 Feb 06]. Available from: http://www.fao.org/faostat/en/#data/
26. Sunandini GP, Paul KSR, Irugu SD. Analysis of trend growth and instability in rice production in Andhra Pradesh. Curr J Appl Sci Technol. 2020;39(42):40–46. https://doi.org/10.9734/cjast/2020/v39i4231129
27. Nkwabi J, Sharma R, Dev K, Samriti and Sharma, S. Challenges for small scale rice farmers – A case study from Tanzania. Econ Aff. 2021;66(1):149–155. https://doi.org/10.46852/0424-2513.1.2021.19
28. Suleiman R. Local and regional variations in conditions for agriculture and food security in Tanzania: A review. AgriFoSe2030 report; 2018.
29. Wineman A, Jayne TS, Modamba EI, Kray H. The changing face of agriculture in Tanzania: Indicators of transformation. Dev Policy Rev. 2020;38(6):685–709. https://doi.org/10.1111/dpr.12491

AUTHORS:
Bathobile Maseko[1]
Gregor Feig[2]
Roelof Burger[3]

AFFILIATIONS:
[1]South African Weather Service, Pretoria, South Africa
[2]South African Environmental Observation Network, Pretoria, South Africa
[3]Unit for Environmental Sciences and Management, North-West University, Potchefstroom, South Africa

CORRESPONDENCE TO:
Bathobile Maseko

EMAIL:
bathobile.maseko@weathersa.co.za

DATES:
Received: 06 Mar. 2020
Revised: 12 Feb. 2021
Accepted: 29 Mar. 2021
Published: 29 Sep. 2021

HOW TO CITE:
Maseko B, Feig G, Burger R. Estimating lightning NO_x production over South Africa. S Afr J Sci. 2021;117(9/10), Art. #8035. https://doi.org/10.17159/sajs.2021/8035

ARTICLE INCLUDES:
☒ Peer review
☐ Supplementary material

DATA AVAILABILITY:
☐ Open data set
☐ All data included
☒ On request from author(s)
☐ Not available
☐ Not applicable

EDITOR:
Yali Woyessa

KEYWORDS:
Lightning Detection Network, lightning flash density, nitrogen oxides, EDGAR, lightning NO_x

FUNDING:
None

Estimating lightning NO_x production over South Africa

Nitrogen oxides (NO_x = NO + NO_2) are toxic air pollutants and play a significant role in tropospheric chemistry. Global NO_x hotspots are the industrialised regions of the USA, Europe, Middle East, East Asia and eastern parts of South Africa. Lightning is one of the many natural and anthropogenic sources of NO_x to the troposphere. It plays a role in the formation of particulate matter and tropospheric ozone, which are both linked to harmful health and climate effects. The discourse on NO_x over the southern African continent has mainly focused on anthropogenic sources. However, lightning is known to be a main source of tropospheric NO_x globally. It is therefore important to understand its contribution to the national and global NO_x budget. Data from the South African Lightning Detection Network were used to approximate the influence of lightning on the NO_x load over the country, and to develop a gridded data set of lightning-produced NO_x (LNO_x) emissions for the period 2008–2015. The Network monitors cloud-to-ground lightning strikes; and theoretically has a detection efficiency of 90% and a location accuracy of 0.5 km. An emission factor of 11.5 kg NO_2/flash was employed to calculate the LNO_x budget of ~270 kt NO_2/year. The calculated LNO_x was 14% of the total NO_x emission estimates published in the EDGAR v4.2 data set for the year 2008. The LNO_x emission inventory will improve model performance and prediction, and enhance the understanding of the contribution of lightning to ambient NO_2.

Significance:

- The results show that both lightning and industrial NO_x sources are essential in evaluating NO_x and tropospheric O_3 chemistry over South Africa. As a result they should both be considered in air quality modelling, to assist in air quality management planning.
- LNO_x emissions are projected to increase with climate change, which may lead to an increase in tropospheric O_3. Thus it is important to have an LNO_x inventory, to be used as input into air quality modelling, as it will improve model performance and forecasting, and the understanding of the sensitivity of ambient pollution to changes in lightning emission.
- It will further inform chemical transport modelling so that the contribution of both natural and anthropogenic sources can be better understood.

Introduction

Poor air quality is a key environmental concern in South Africa, as it poses a serious threat to the well-being of the people of South Africa. Two of the key pollutants with adverse health and environmental impacts are nitrogen oxides (NO_x= nitric oxide (NO) + nitrogen dioxide (NO_2)) and particulate matter (PM). NO_x have effects that are felt on humans and the environment, but they are also important reagents in tropospheric chemical processes that result in the creation of secondary atmospheric pollutants, like ozone (O_3).

To alleviate the effects of air pollution, it is essential to have a comprehensive understanding of the sources of atmospheric pollutants to manage emissions. Management of air quality requires that the quantity of pollutants released into the atmosphere from sources be known, to determine how much of the emissions need to be reduced to achieve acceptable levels. Furthermore, the understanding of pollutant sources is essential for atmospheric chemical transport modelling, which is a valuable tool in comprehending the distribution of the pollutants and their potential impacts.[1]

Areas that are strongly affected by poor air quality in South Africa include the Mpumalanga Highveld, Vaal Triangle area, and the Waterberg Bojanala area. These areas have been affirmed as air quality priority areas in terms of Section 19 of the *National Environmental Management: Air Quality Act*.[2] The primary sources of pollution in these areas are coal-fired power plants, vehicles, domestic fuel burning, and metallurgical and petrochemical industries.[3,4]

Carbon monoxide (CO), ozone (O_3), particulate matter (PM_{10} and $PM_{2.5}$), nitrogen dioxide (NO_2) and sulfur dioxide (SO_2) are pollutants that are listed as being of particular concern in South Africa.[5] These criteria pollutants are regulated by the National Ambient Air Quality Standards, as they have adverse health and environmental impacts.[6] Nitrogen oxides are a focus of this study and are important in many ways:

- Long-term exposure to NO_2 may reduce the functionality of the lungs.[7]
- NO_2 increases the risk for children and the elderly of acquiring diseases such as bronchitis.[8]
- NO_x is vital in climate and atmospheric chemistry; it reacts with volatile organic compounds in the presence of sunlight to form O_3,[9] which in the troposphere has adverse health impacts and is a greenhouse gas[10].

- NO_x is a precursor for the formation of secondary atmospheric aerosol, and it is involved in the formation of pollutants implicated in acid deposition.[11]

Satellite analysis of total NO_2 vertical column density has identified some NO_x hotspots over the industrialised areas of the USA, Europe, Middle East, East Asia and South Africa (Figure 1).[12] South Africa's hotspot is over the northeastern parts of the country (the Highveld). This area contributes 90% of the industrial NO_x emissions (Figure 2).[13] The Mpumalanga Province of South Africa is one of the world's largest NO_2 hotspots, with 12 coal-fired power plants situated in the area.[14] Other sources of NO_x in the Highveld region include petrochemical industries (particularly coal and gas to liquid facilities), metallurgical smelters, road transport, biomass burning, and human settlements.[13] Approximately 8 million people in the Highveld priority area and neighbouring Gauteng conurbation are exposed to poor air quality.[14]

Nitrogen oxides are released into the atmosphere from anthropogenic and natural sources. Anthropogenic sources are industrial activities, fossil fuel combustion, transportation and power plants. Natural sources include lightning and soil.[15] Globally, 60–70% of the total NO_x budget is from anthropogenic sources rather than natural sources.[16] Therefore, natural sources of NO_x in general, and lightning in particular, have received less attention. This is mainly due to the high spatiotemporal variability associated with the detection of lightning.[17] Whilst this remains the case, lightning-produced NO_x (LNO_x) comprises about 10–15% of the global NO_x budget and is a leading source of NO_x in the upper troposphere.[18] Consideration of the dynamics of LNO_x production is essential in understanding atmospheric chemical processes and the ambient concentrations of NO_x and O_3.[19]

Tropospheric O_3 concentration depends on the imminent precursor emissions, together with the changes in meteorological variables, including temperature and atmospheric moisture.[20] Various studies have suggested that future lightning activity is projected to intensify due to climate change. At the upper troposphere, LNO_x is more efficient in producing tropospheric O_3; thus, developing a regional LNO_x emission inventory will assist in understanding the sensitivity of the increase in lightning activity, with climate change, to O_3 in the troposphere.[21-23]

During a thunderstorm, LNO_x is carried to the upper troposphere by convective updrafts, where it is more effective at generating O_3 and has a longer lifespan than in the lower troposphere, where most of the anthropogenic NO_x is released.[24] Thus, to be able to provide a reliable tropospheric O_3 budget, determination of an accurate LNO_x load is vital.[19] Estimation of LNO_x is a challenge compared to that of anthropogenic emissions because the occurrence of lightning varies significantly in space and time.[17] LNO_x production estimates range within the [1–20] Tg (N) per year, and most estimates from different studies point toward the lower end of the production range of ~ 5 Tg (N)/year.[25,26] Although the production points towards the lower end, the evidence is not enough to reject the upper end.[19,27,28]

In estimating LNO_x, a bottom-up approach is commonly used to measure (1) NO_x production per energy unit, (2) energy discharged per flash and (3) flash frequency, and to estimate LNO_x as the product of these quantities.[17] Values cited in the literature range in magnitude, due to various assumptions and laboratory measurements. Another complication arises from the differences in cloud-to-ground (CG) and intra-cloud (IC) flashes. Previous studies suggested that IC flashes are more recurrent but less energetic than CG flashes (30% of total lightning), therefore less effective as NO_x producers compared to CG flashes (IC produces 10% of the NO_x produced by CG per flash).[17,25,29]

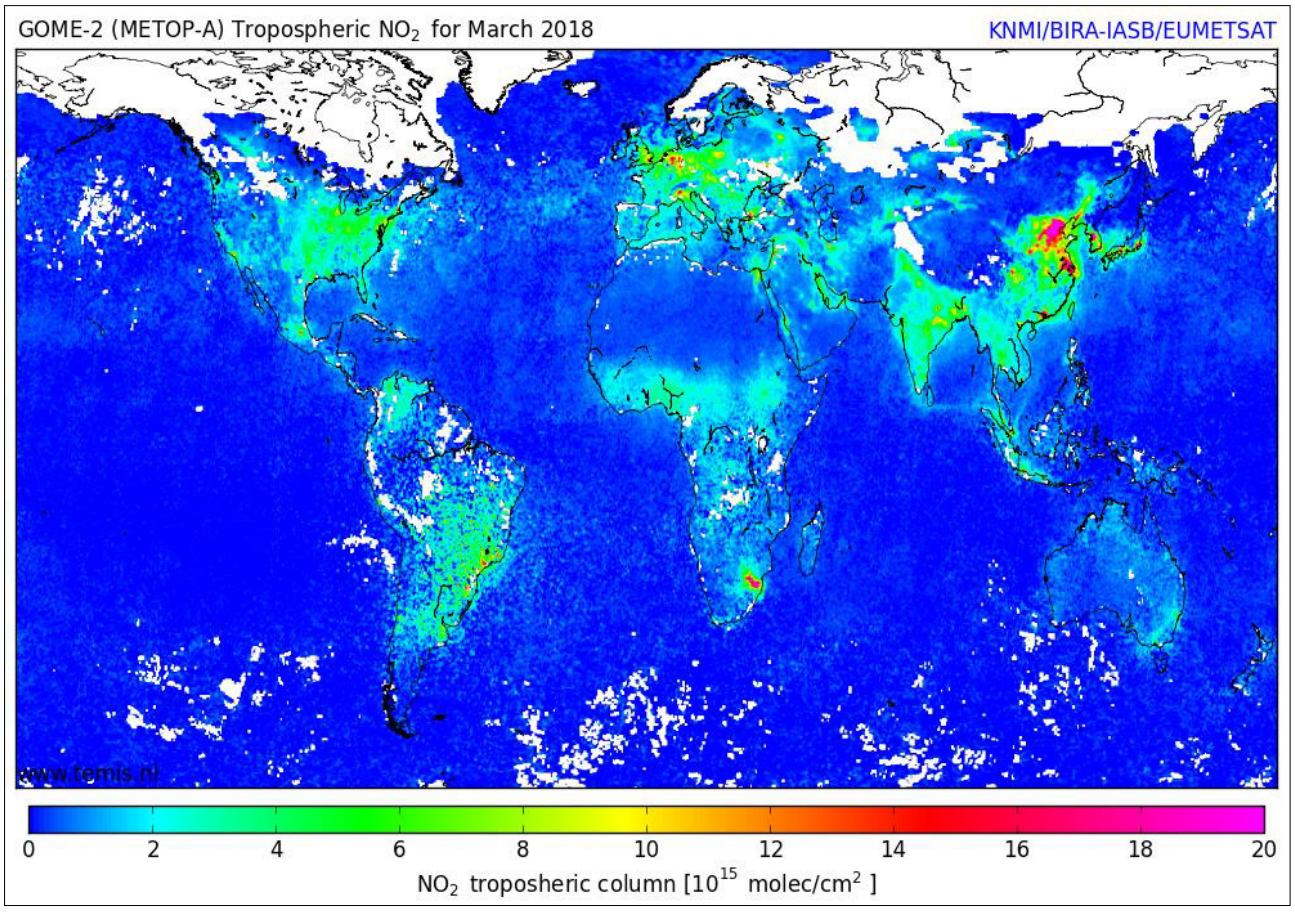


Figure 1: Tropospheric NO_2 column density in 10^{15} molecules/cm^2 for March 2018.[12]

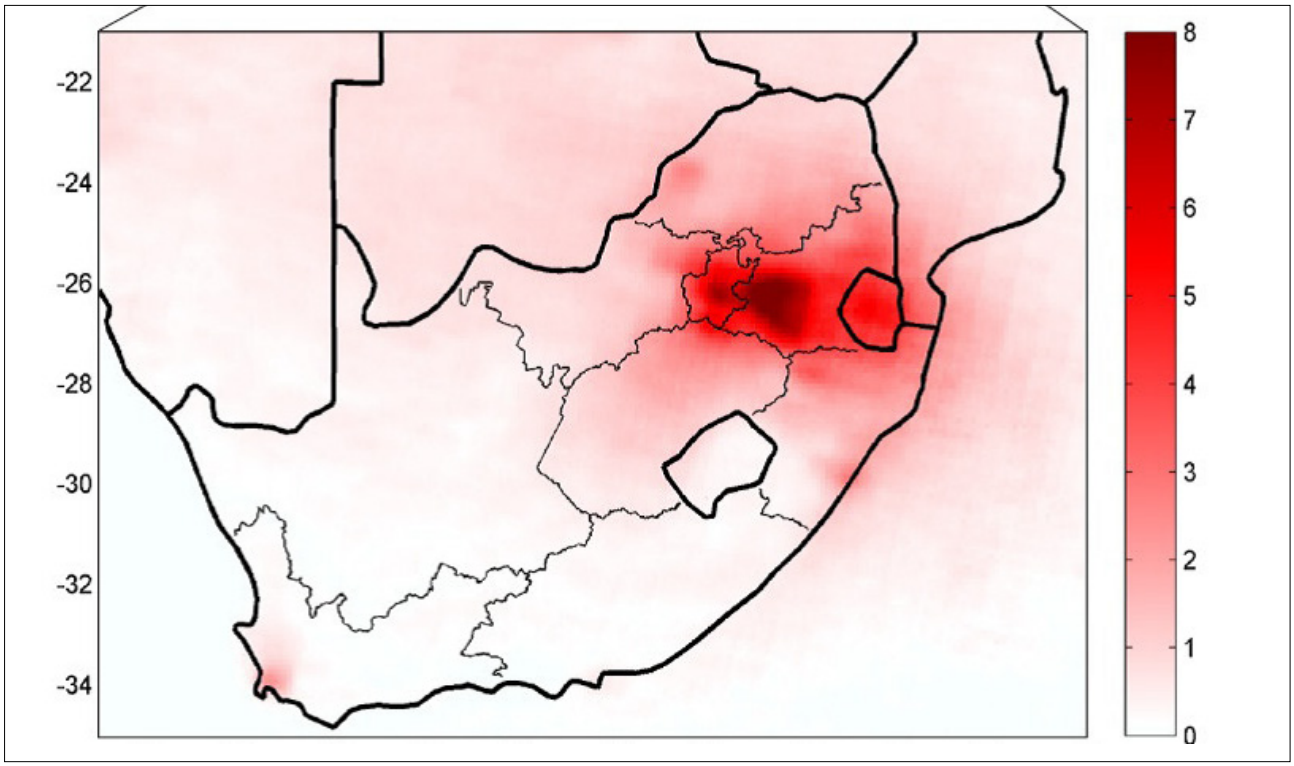


Figure 2: Tropospheric vertical NO_2 column over the Highveld of South Africa.[13]

However, recent studies have proposed that IC flashes may dissipate similar energy to that of CG flashes, and thus IC flashes may produce the same amount of NO_x as CG flashes.[24,30-33] This is contradictory to the common assumption that IC flashes produce less NO_x than CG flashes, and highlights that, despite research efforts, there is still a great deal of uncertainty that remains regarding NO_x production on a per flash basis, as well as the relative production by IC and CG flashes. Additionally, calculations for LNO_x do not consider the effect that the water vapour has on the production of NO_x. Peyrous and Lapeyre (1982) have shown that NO_x production is strongly dependent on relative humidity.[25]

There is currently limited research regarding the role of the production of NO_x by lightning in regional atmospheric chemistry. Of the studies that exist, most concentrate on the global LNO_x production.[23,34,35] The global studies of LNO_x might not be applicable at the regional scale, because they may not capture the number of lightning strikes occurring within a region, the variability involving the production of NO_x per flash, and the varying ratio of IC and CG flashes.

This study builds on the work that has been done by Ojelede et al.[36] They conducted a study to estimate LNO_x production over the Highveld of South Africa for the year 2002, utilising lightning data obtained from the Lightning Position and Tracking System (LPATS) network. The LPATS has a detection efficiency of 80% and consists of six sensors over the eastern half of the country. The study was done on a limited spatial extent, for a short study period.

In this study, we used lightning data from the South African Lightning Detection Network (SALDN) that is operated by the South African Weather Service. It has coverage over the whole country, with a 90% detection efficiency for CG lightning and location accuracy of 0.5 km, i.e. the SALDN can detect at least 90% of all CG flashes and position them within 0.5 km. Information from SALDN allows for the high-resolution identification of areas of intense lightning occurrence. It also provides an opportunity to investigate the spatiotemporal characteristics of LNO_x production, which can then be used as an input into atmospheric chemical transport modelling and to improve air quality modelling. The data can also be used for validation of model simulations.

Data and methods

Lightning data from the SALDN were used in this research. The South African Weather Service installed the SALDN in 2005. Initially, it consisted of 19 sensors across South Africa but has since been expanded to 25 sensors including the one in eSwatini (Figure 3). The data are available from 2006 until the present; however, for this study, 8 years of data were used (2008–2015). The SALDN has a 90% detection efficiency of CG lightning and location accuracy of 0.5 km.[37] There have been some changes since the sensors were operationalised in 2006. Sensors have been added, and others have been relocated to ensure optimal network performance and to overcome some environmental developments and noise sources near some of the sensors. The network expansion and relocation of some sensors in 2009/2010, 2011, as well as in 2015, have greatly enhanced the accuracy of the lightning detection network; the accuracy is based on the South African Weather Service in-house quality control software.

The South African Weather Service network uses Vaisala instruments including two types of sensors: LS7000 and LS7001 model sensors. The LS7000 sensors were installed prior to 2009, and the LS7001 sensors were installed from 2009 onwards. These sensors measure the electromagnetic signature of a lightning flash using a low-frequency bandwidth because the radiation from CG strokes is strongest within that band.[38] As only a small percentage of IC flashes are measured at that bandwidth, the network sensors are primarily used to locate CG flashes.[38]

The methods used to determine the position of individual CG strokes are (1) magnetic direction finding (MDF) and (2) time of arrival (TOA).[39] The combination of MDF and TOA techniques offers high detection efficiency and accurate location detection for CG lightning strokes from only two sensors. When the MDF and TOA methods are used individually, they require three or more sensors to detect lightning. It is impossible to get a 100% accurate position of lightning ground stroke even with combined MDF and TOA techniques. However, a degree of accuracy can be determined. For a thorough description of the method see Gill[39].

The lightning data retrieved using the Fault Analysis and Lightning Location System, contain information such as date and time, latitude,

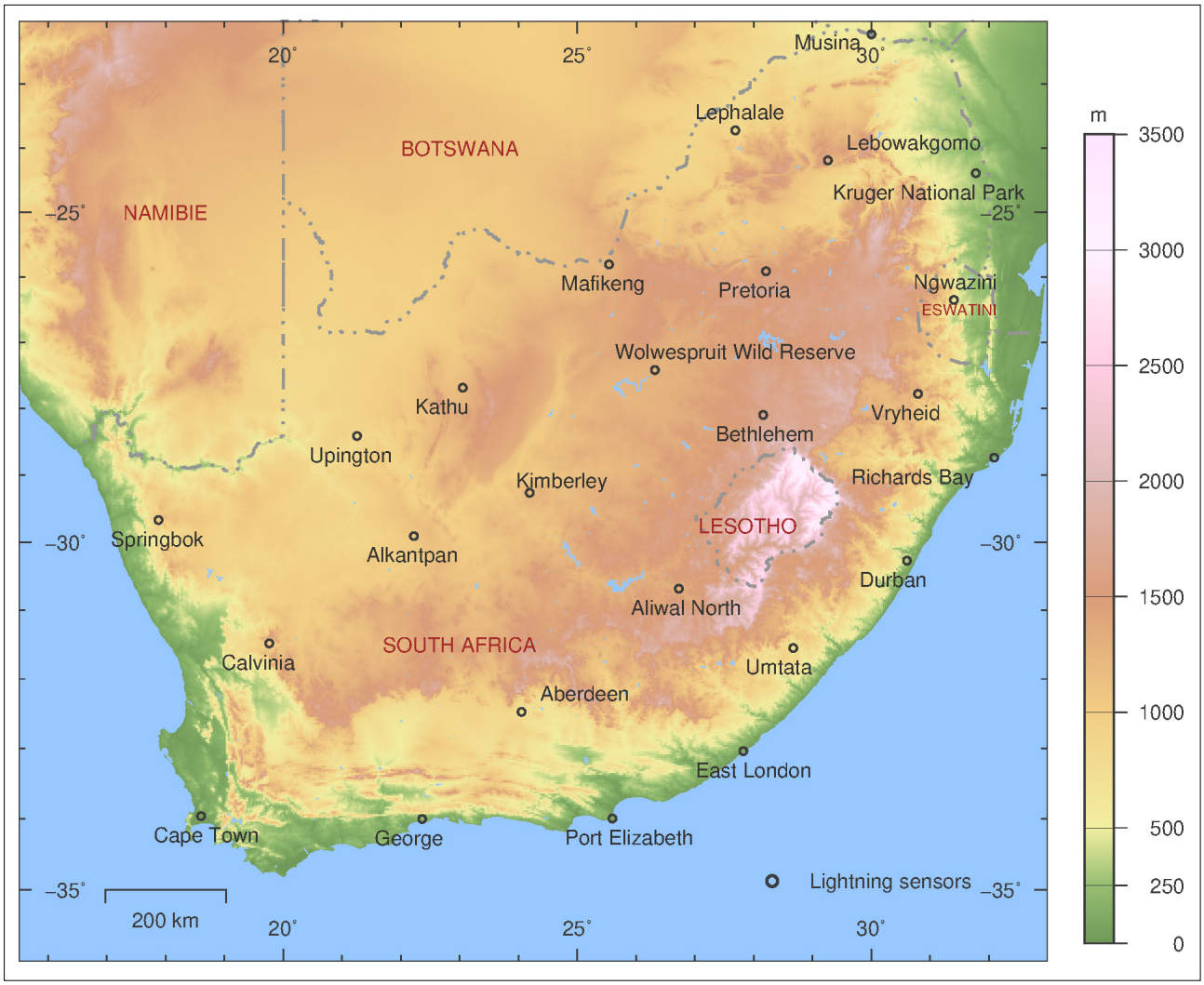


Figure 3: Position of the 25 Vaisala lightning detection sensors over South Africa

longitude, and peak ampere. All CG lightning flash data were taken into consideration, regardless of polarity. Annual lightning ground flash density was calculated on $0.1^o \times 0.1^o$ grid boxes over the country and is expressed in flashes/km^2. The number of lightning flashes that were recorded by the SALDN over the 8 years (2008–2015) was counted for each individual grid box over the country. The sum of lightning flashes was divided by the area of the grid box to give the number of lightning flashes per square kilometre. To get the average annual lightning flash, the number of lightning flashes per square kilometre was divided by 8, as 8 years of data were considered. All grid boxes were considered, irrespective of the number of flashes. The Grid Analysis and Display System was used to display the lightning ground flash density map.

Information on total flash rate and production per flash is needed to quantify LNO_x production. NO_x emission factors from the literature and CG lightning flash data from the SALDN were used. In South Africa, there have been no studies done to estimate the IC/CG ratio. Therefore, IC/CG ratio estimates from other countries were used for South Africa (Table 1). Lightning characteristics – the number of strokes per flash, channel length, and ratios of IC and CG flashes – can vary.[34] Furthermore, the IC/CG flash ratio varies intensely during the life cycle of a thunderstorm. Therefore, it is assumed that such a ratio is a rough estimate as it is influenced by many factors, which include the severity and phase of a storm. Many studies have proposed a global annual mean IC/CG flash ratio of approximately 2 to 4.[40] We used a IC/CG flash ratio of 3 – the most often used ratio for the mid-latitude to sub-tropic regions.[25,41]

Various emission factors from different studies have been used (Table 2). We used the emission factor from Schumann and Huntrieser[40] for this study. Beirle et al.[35] referred to it as being the currently best-accepted emission factor. Using the emission factor from Schumann and Huntrieser[40], one should bear in mind that global LNO_x production might not be entirely applicable to regional LNO_x production. Schumann and Huntrieser[40] concluded that the most likely value of global LNO_x production is 5 Tg N/year with an uncertainty of 2–8 Tg N/year. The global annual mean flash frequency is 44±5/second, and a mean production per flash is 250 moles NO_x/flash (which is approximately 11.5 kg NO_2/flash).

Based on previous studies[24,32,33,47], we assumed that IC flashes generate the same amount of NO_x as CG flashes.Total LNO_x (kt NO_2/year) to total flashes was calculated using Equation 1:

$$Total\ (LNO_x) = P(NO_2)(IC) + P(NO_2)(CG) \qquad \text{Equation 1}$$

where

- $P(NO_2)$ is the production of NO_2 = 11.5 kg NO_2/flash
- IC = 3 x CG flashes
- CG = CG flashes

Annual LNO_x maps were generated to be able to observe the distribution of LNO_x over South Africa. Equation 1 was used to calculate the total LNO_x on a $0.1^o \times 0.1^o$ resolution over the country. The lightning flash information from the SALDN, the emission factor from Schumann and Huntrieser[40], and the IC/CG ratio of 3 were used for calculations.

Table 1: Intra-cloud to cloud-to-ground lightning flash ratio from different literature

IC/CG flash number ratio	Reference	Location
3.35 (2 to 6)	Prentice and Mackerras[42]	Brisbane, Australia
3	Price et al.[25]	USA
4.4±1	Mackerras et al.[43]	Brisbane, Australia
3.5	Nesbitt et al.[44]	USA
Over 100	Dye et al.[45]	Northeastern Colorado, USA
2.4	Bond et al.[46]	Tucson, Arizona, USA
2.8±1.4 (1 to 9)	Boccippio et al.[41]	Continental USA

IC, intra-cloud; CG, cloud-to-ground

Table 2: Estimates of LNO_x production per flash from different studies

Production rate (kg NO_2/flash)	Method	Reference
3.5 (1.2 to 10.6)	GOME and NLDN	Beirle et al.[49]
51.2	Theoretical	Price et al.[25]
16.4	3D cloud model, in-situ measurements, EULINOX	Fehr et al.[50]
4.1 (1.5 to 11.0)	GOME and NLDN	Beirle et al.[51]
11.5 (1.5 to 30.5)	Review	Schumann and Huntrieser[40]
9.2 to 23.0	STERAO-A, ONERA, NLDN	DeCaria et al.[32]
23	3D cloud model	Ott et al.[33]
4.6 to 11.5	OMI	Bucsela et al.[52]
11.4 (1.5 to 30.5)	SCIAMACHY	Beirle et al.[35]

The total LNO_x emission is expressed in kg NO_2/km^2/year. A box plot was created using the R-software lattice package to observe how LNO_x production varies from month to month.[48] The procedure for calculating LNO_x is similar to the one described above, expressed in kt NO_2/month. To produce the LNO_x diurnal variation graph, we used hourly lightning flash data for the 8-year period. The calculations were made over the 24-h period, considering every hour, from 00 hour to 23 hour, e.g. lightning flash considered for 00 hour is the one greater or equal to 00 hour but less than 01 hour.

Results and discussions

Figure 4 indicates the annual ground flash density for the entire country, showing how the lightning flash density is distributed over South Africa. The results show that high flash densities occur over Mpumalanga, KwaZulu-Natal and Gauteng Provinces, with the highest flash density over the far northern parts of KwaZulu-Natal and the Mpumalanga escarpment. The flash density declines towards the west of the country, along the coast and towards the northeastern parts of the country. Over the mountainous region of Lesotho, flash densities of ≤5 flashes/km^2 were observed. Gijben[37] and Gill[39] presented similar results.

Lightning is associated with convective storm development; hence portions of the country that experience many convective storms – the central, eastern and northern parts of the country – record high numbers of lightning flashes. The western and southwestern parts of the country are almost without thunderstorms, and therefore experience less lightning.[39] This is due to the influence of topography in enhancing thunderstorm development, and therefore increasing lightning activity.[53] Various studies internationally have shown a positive relationship between topography and lightning.[54-56] The decrease in lightning over the elevated areas of Lesotho (Figure 4) can be attributed to the reduction in the occurrence of lightning at altitudes above 2000 m.[37,53]

Figure 5 shows the spatial distribution of LNO_x over South Africa for the 8-year period (2008–2015). Areas of high lightning flash density (Figure 4) are regions that incur high LNO_x production. Highly elevated areas have a high production of NO_x, particularly over the northeastern parts of the country, except over the elevated areas of Lesotho where lightning occurrence is reduced because of altitudes above 2000 m. The Highveld region of South Africa is one of the regions where lightning is frequent and therefore has a high LNO_x production rate. This is also a region which is heavily industrialised and is the source of a large proportion of anthropogenic NO_x production.[57] In previous studies, in which LNO_x was not considered, it was found that the leading producer of NO_x in the Highveld Priority Area was industrial sources (power generation, petrochemicals and metallurgical industries) which are responsible for almost 96% of NO_x emissions.[58,59]

Production of NO_x by lightning differs seasonally, in accord with lightning distribution. High production of LNO_x occurs in summer (NDJ) and autumn (FMA). This high LNO_x production is associated with deep convective thunderstorms, which are accompanied by lightning. During spring, a moderate amount of lightning occurs, while in winter there is close to zero lightning occurrence as there is lack of moisture and the atmosphere is stable, hence production is at a minimum during winter (MJJ).

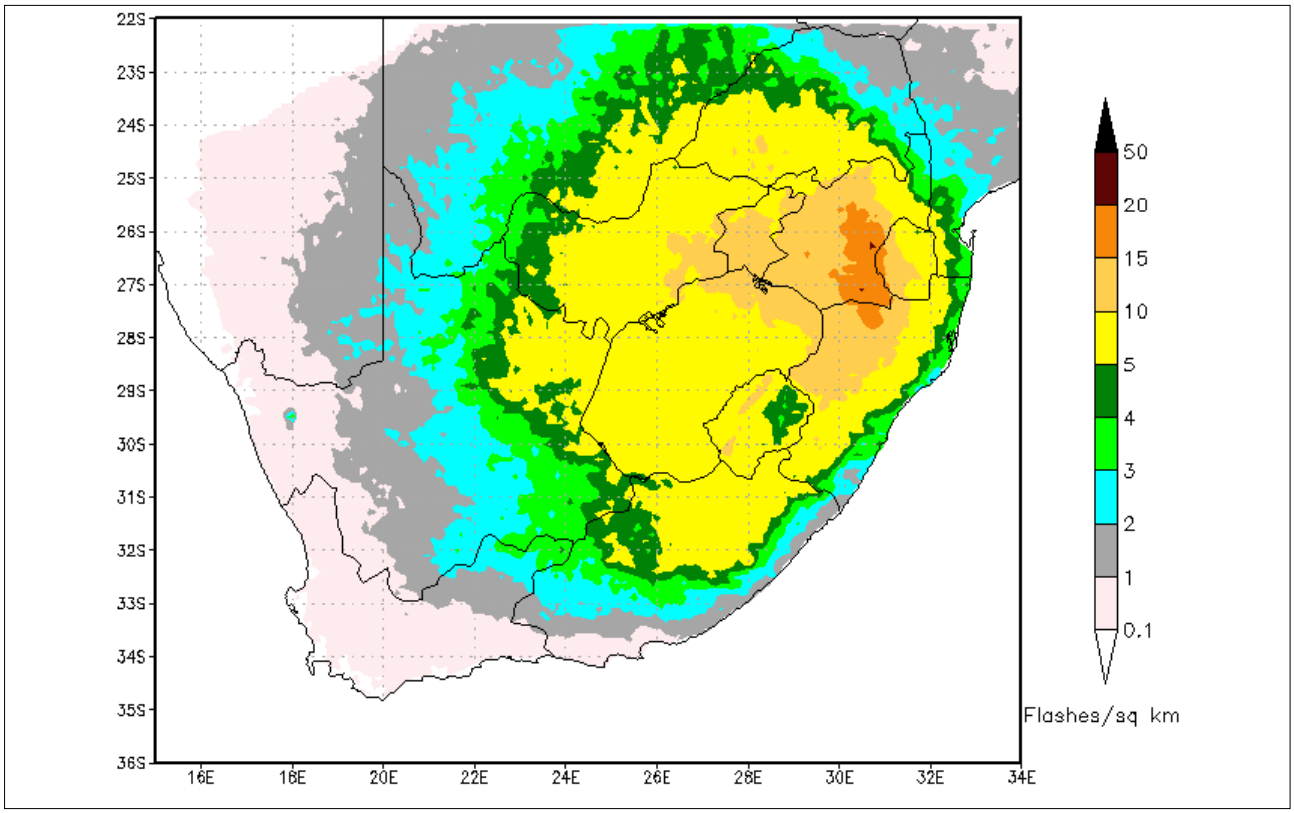


Figure 4: Average number of lightning ground flashes per km² per year over South Africa for the 8-year period (2008–2015).

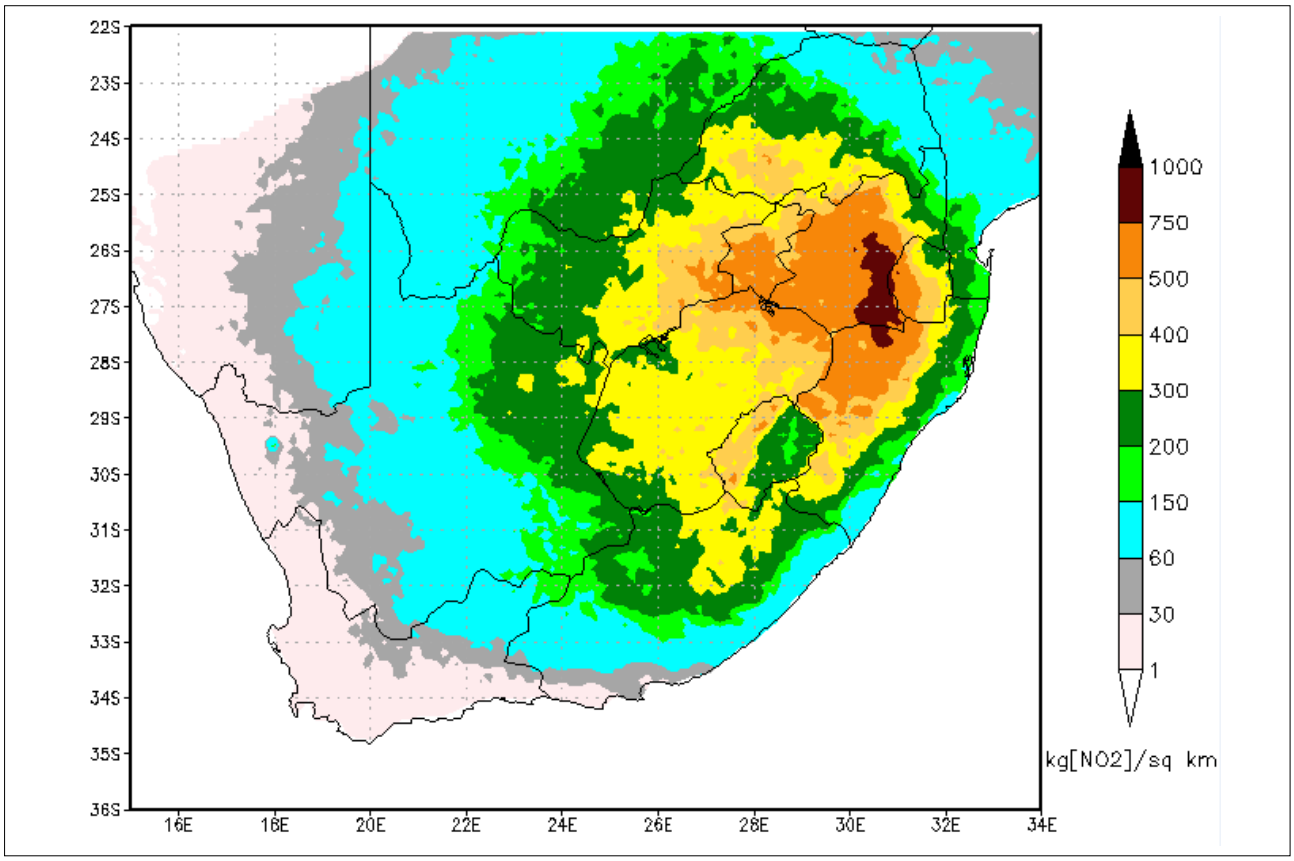


Figure 5: Average number of total LNO_x in (kg NO_2/km²/year), using the emission factor of Schumann and Huntrieser[40].

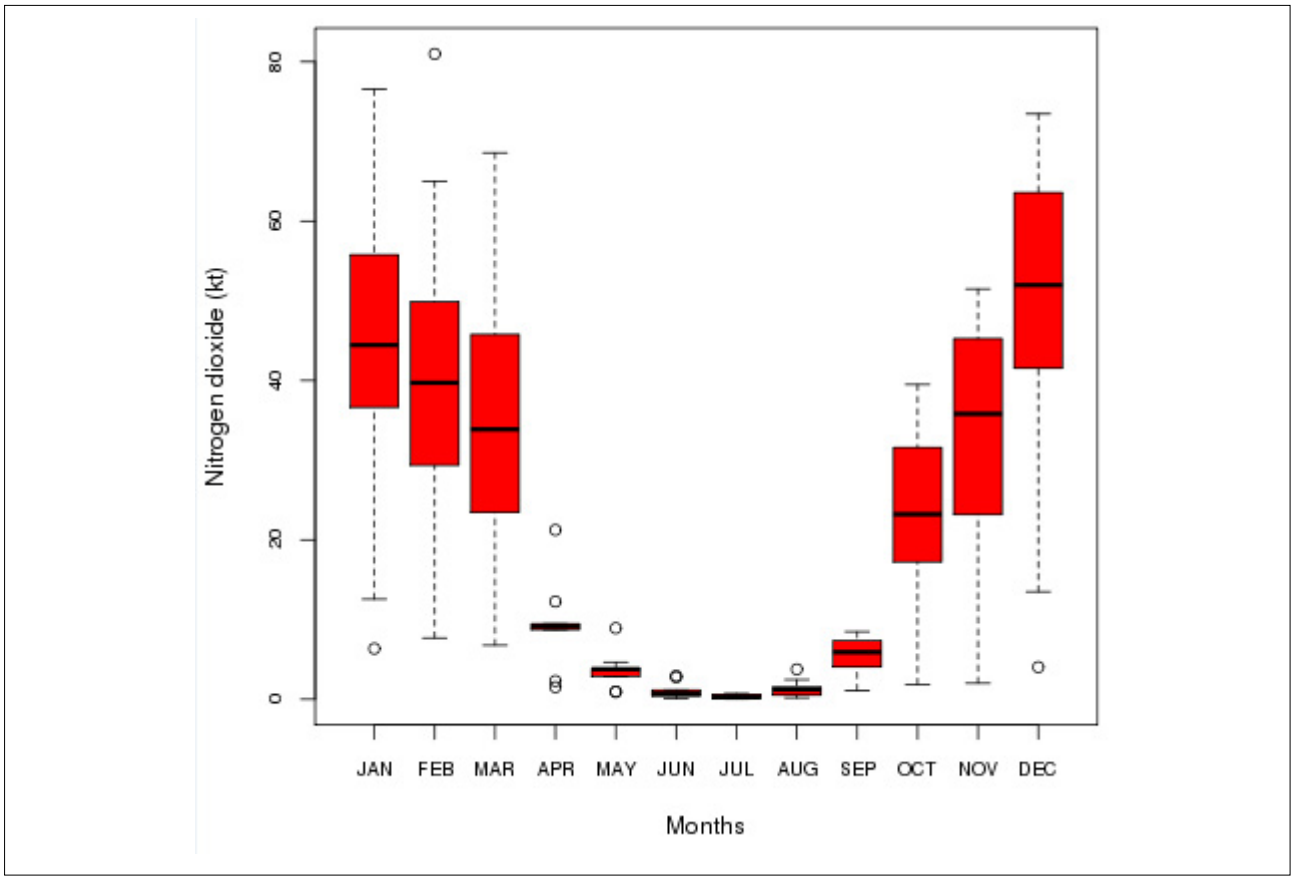


Figure 6: Monthly estimates of LNO_X over South Africa.

Table 3 shows the total LNO_x generated, accompanied by the standard deviation of the interannual variability. The total LNO_x produced is estimated to be ~ 270.85 (±42.5) kt NO_2/year, with summer contributing 51%, autumn 35%, spring 12%, and winter 2% of the annual budget. Figure 6 shows the variation of lightning production of NO_x from month to month, with the box-and-whisker plot exhibiting the minimum and maximum LNO_x production values, as well as the interquartile range and median. The lowest LNO_x production is noticeable in winter months and the highest production in summer months. The median values depict that LNO_x production is higher for December, followed by January and February. LNO_x production further indicates relatively high values during spring and autumn as well, specifically in November, March and October.

Table 3: Seasonal and annual LNO_x production over South Africa excluding Lesotho and eSwatini

Season	Total LNO_x production (kt)
Summer	137 (± 27)
Autumn	94 (± 21)
Winter	6 (± 3)
Spring	34 (± 11)
Total LNO_x	270.85 (kt NO_2/year)

The seasonal LNO_x estimates presented depict a similar pattern to that of Ojelede et al.'s[36] with summer contributing more to LNO_x annual estimates, followed by autumn, spring, and then winter. The total annual LNO_x production obtained by Ojelede et al.[36] was 24% of that recorded in this study. The 65 kt NO_2/year obtained is for when IC flashes are assumed to produce only 10% of NO_x produced by CG flashes. Not assuming equal production of IC and CG flashes can lead to underestimation of LNO_x.[36]

The diurnal variation of LNO_x in Figure 7 shows low production of NO_2 in the early morning, with peaks starting from ~0900 local time, reaching a maximum in the late afternoon around ~1500 local time before decreasing. The late afternoon peak in LNO_x may be because of the surface being heated due to the incoming solar radiation during the day, which then induces convective activity locally. Collier et al.[60] indicated the peak of annual lightning activity to be around 1700 local time in Pretoria, South Africa. Confirmation of these diurnal emission cycles improves the efforts to model LNO_x emissions over the country. The diurnal variation of LNO_x production is similar for summer, autumn and spring with LNO_x reaching its peak in the late afternoon, except during winter (Figure 8). During winter (green), there is not much variation in LNO_x production; it remains more or less the same throughout the day, and it is lower than during other seasons.

The calculation of the annual LNO_x production allows us to compare this particular source of NO_x to the South African atmosphere with the better known and understood industrial and vehicular sources, in order to determine the importance of LNO_x in the South African NO_x budget. Figure 9 shows NO_x emission from lightning, together with NO_2 emission obtained from the Emission Database for Global Atmospheric Research (EDGAR v4.2) for South Africa, which includes anthropogenic and biomass burning emissions.

The EDGAR v4.2 data used in the study provide estimates of the global biomass burning and anthropogenic emissions. The existing data are for all countries, with emissions given per main source category, on a spatial resolution of 0.1° x 0.1° grid over the globe. The database was obtained from the European Commission's Joint Research Centre data catalogue for the entire South Africa.[61] According to EDGAR v4.2 data (Table 4), the largest anthropogenic source of NO_x is the power generation sector (1A1a) which accounts for 1078.7 kt of NO_2 per year.

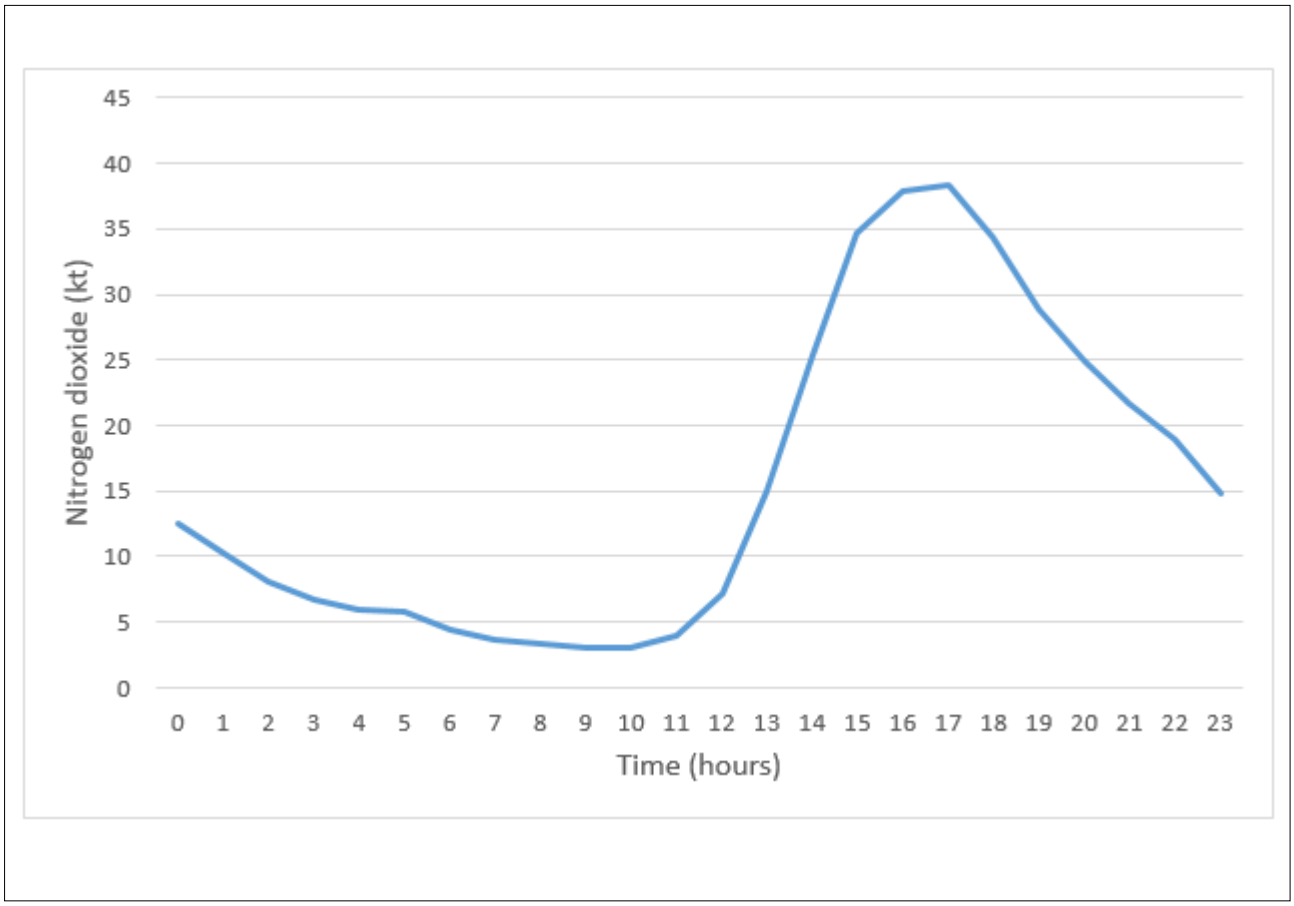


Figure 7: Mean annual diurnal variation of LNO_X production over South Africa for the years 2008 to 2015.

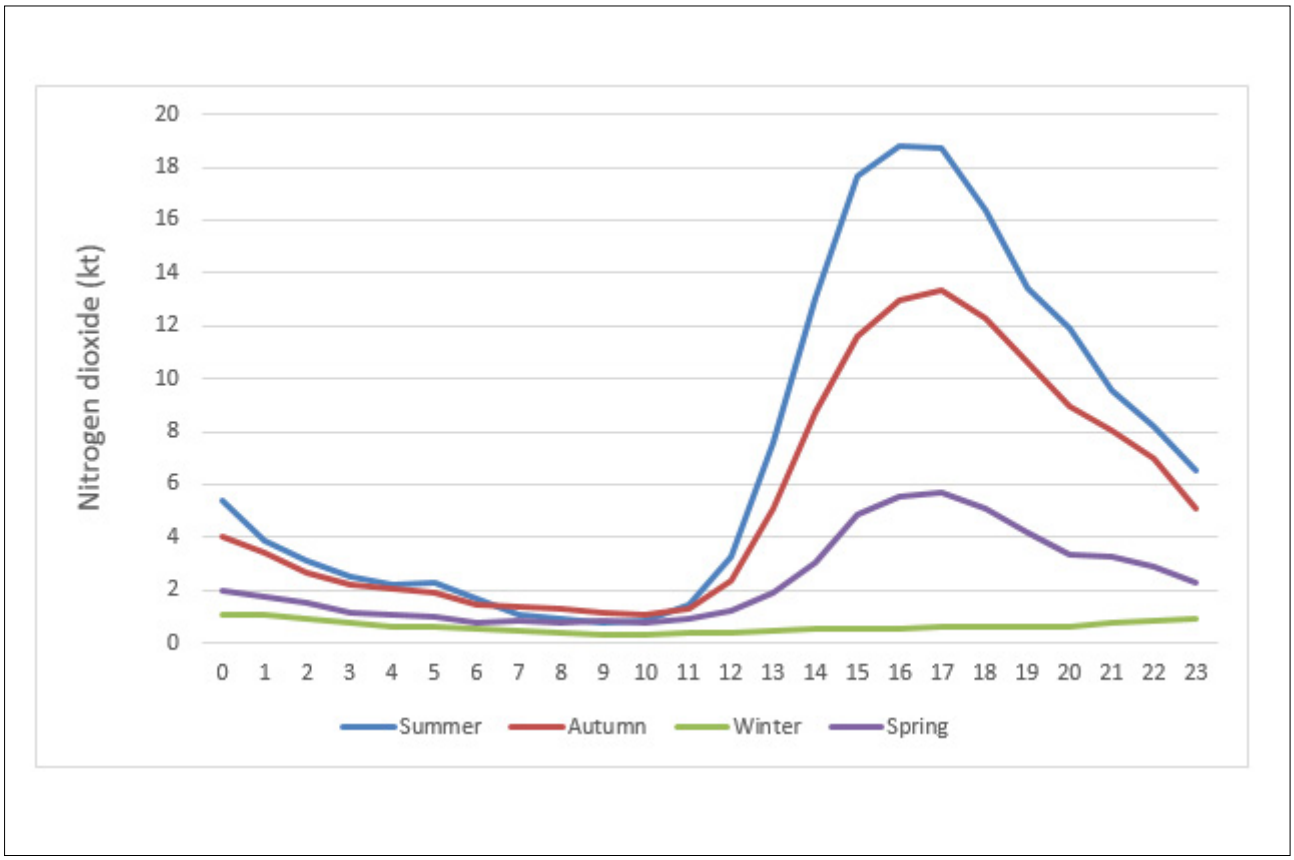


Figure 8: Seasonal diurnal variation of LNO_X production over South Africa.

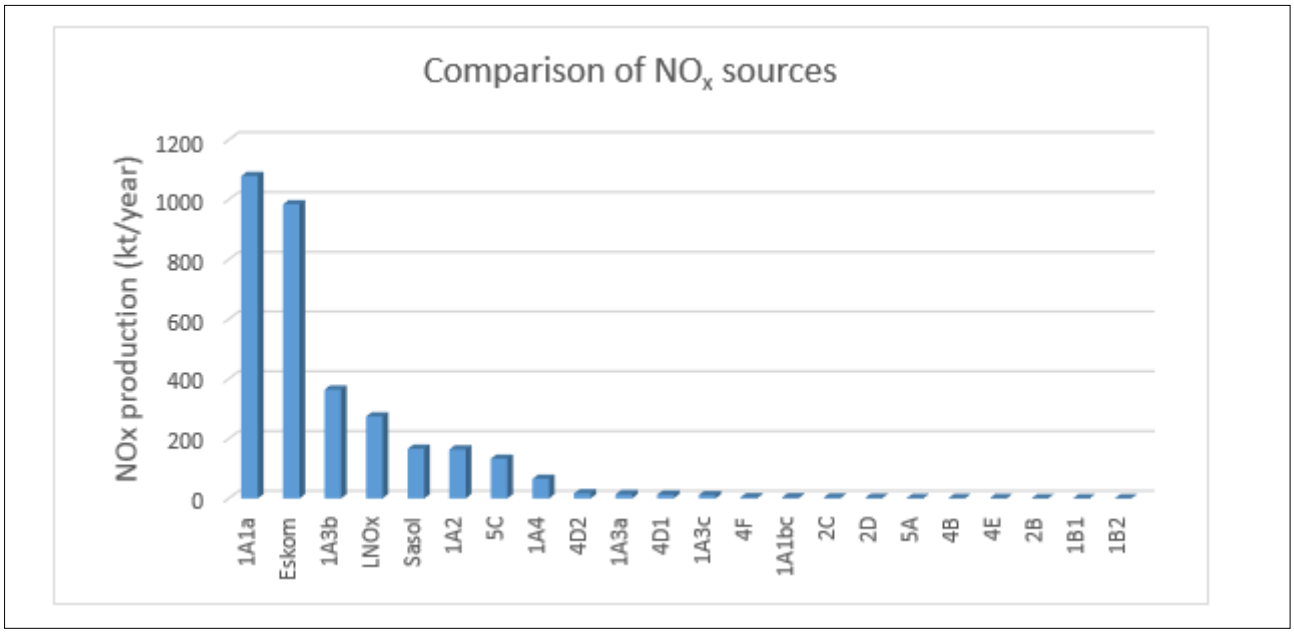


Figure 9: NO_2 emission from EDGAR v4.2 and lightning for the year 2008 over South Africa, in kt NO_2/year.

Table 4: NO_2 emission from EDGAR v4.2 and LNO_x for the year 2008 over South Africa

Sectors (code)	NO_x production (kt NO_2/year)
Public electricity and heat production (1A1a)	1078.7 (57.5%)
Road transportation (1A3b)	364.3 (19.4%)
Lightning NO_x	270.9 (14.4%)
Manufacturing industries and construction (1A2)	163.7 (8.7%)
Grassland fires (5C)	132.6 (7.1%)
Residential and other sectors (1A4)	65.4 (3.5%)
Manure in pasture/range/paddock (4D2)	17.1 (0.9%)
Domestic aviation (1A3a)	14.2 (0.8%)
Direct soil emissions (4D1)	12.5 (0.7%)
Rail transportation (1A3c)	10.7 (0.6%)
Agricultural waste burning (4F)	3.8 (0.2%)
Other energy industries (1A1bc)	3.2 (0.2%)
Production of metals (2C)	2.6 (0.1%)
Production of pulp/paper/food/drink (2D)	2.1 (0.1%)
Forest fires (5A)	1.2 (0.1%)
Manure management (4B)	0.9 (0.05%)
Savanna burning (4E)	0.8 (0.04%)
Production of chemicals (2B)	0.7 (0.03%)
Fugitive emissions from solid fuels (1B1)	0.7 (0.03%)
Fugitive emissions from oil and gas (1B2)	0.1 (0.01%)

Road transportation (1A3b) is the second largest emitter of NO_x, emitting 365.3 kt of NO_2 over South Africa, accounting for 19.4% of the total NO_x from EDGAR v4.2. Road travel dominates the transport sector, with the vehicle/ownership ratio in South Africa higher than the world average.[62] This is due to the extensive travel distances between settlements and workplaces.[62] Lightning NO_x production is the third largest source of NO_x to the atmosphere in South Africa and accounts for ~ 270 kt NO_2/year, which is 14% of the total NO_x emitted as estimated in EDGAR v4.2. This percentage indicates that lightning is a significant source of NO_x in the atmosphere and cannot be ignored in air quality management planning; it also points to the importance of including lightning NO_x in the NO_x budget of the country.

The manufacturing industries and construction sector (1A2) is the fourth highest emitter of NO_x at 163.7 kt/NO_2/year and accounts for 8.7% of the total NO_x. Grassland fires (5C) and residential and other sectors (1A4) account for 7% and 3.4%, respectively, of the total NO_x emitted annually. The results indicate that industry is the main source of NO_x and road transportation is the second highest emitter of NO_x. Lightning NO_x production is the third highest emitter of NO_x. Delmas et al.[63] deduced that, at a global scale, combustion of fossil fuels contributes ~50% and biomass burning ~20%, while lightning and microbial activity in soils contribute ~30% of total NO_x emission. A recent study indicated that biogenic NO_x production over the Highveld region amounted to 28.25 kt/year, which was 3.9% of the total for the region. In the Waterberg, biogenic emissions of NO_x amounted to 3.72 kt/year, which is 2.28% of the total NO_x budget in the area.[64]

Conclusions

We investigated the distribution of lightning over South Africa using 8 years of lightning data obtained from the SALDN. These emission factors gave an estimate of the total annual LNO_x production of 270.85 (± 42.5) kt NO_2/year. The LNO_x estimation in this study builds on the work of Ojelede et al.[36], who used 1 year of lightning data to estimate LNO_x production over the Highveld of South Africa. In this study, LNO_x production for the entire South African region was estimated over the 8 years. The Highveld region is the highest contributor to the total LNO_x production in the country.

The summer season accounts for 51% of LNO_x production, while the autumn season accounts for 35%. A sharp decline was experienced in winter, after which LNO_x production steadily increased in spring when transitioning from winter to summer. The diurnal variation of

LNO_x shows that maximum production of NO_x by lightning occurs in the late afternoons and early evenings, and is correlated with the peak occurrence of convective storms. Summer, autumn and spring all follow a similar pattern in which LNO_x was highest in the late afternoon and decreased in the evening after 20:00 UTC. Winter was the only season during which there was no diurnal variation. The characterisation of both the seasonal and diurnal dynamics of LNO_x production is essential for its use in atmospheric chemical transport modelling and provides suitable input information.

Lightning NO_x production contributed 14% to the national NO_x budget, making it the third highest source, followed by power generation (57.5%) and road transportation (19.4%). These results highlight the importance of including lightning NO_x estimates in the NO_x budget for South Africa in further air quality management planning.

Acknowledgement

We thank the South African Weather Service for providing the lightning data used in this study.

Competing interests

We have no competing interests to declare.

Authors' contributions

B.M. led the writing of the manuscript and calculated the lightning NOx production over South Africa. G.F. and R.B. supervised the study and made contributions that helped improve the original manuscript.

References

1. Monks PS, Archibald AT, Colette A, Cooper O, Coyle M, Derwent R, et al. Tropospheric ozone and its precursors from the urban to the global scale from air quality to short-lived climate forcer. Atmos Chem Phys. 2015;15(15):8889–8973. https://doi.org/10.5194/acp-15-8889-2015
2. National environmental management: Air quality act, Act 39 of 2004, Republic of South Africa.
3. South African Department of Environmental Affairs (DEA). Highveld priority area air quality management plan. Pretoria: DEA; 2011.
4. Venter AD, Vakkari V, Beukes JP, Van Zyl P, Laakso H, Mabaso D, et al. An air quality assessment in the industrialised western Bushveld Igneous Complex, South Africa. S Afr J Sci. 2012;108(9/10), Art. #1059. https://doi.org/10.4102/sajs.v108i9/10.1059
5. South African Department of Environmental Affairs (DEA). National framework for air quality management in the Republic of South Africa. Pretoria: Government Printers; 2013.
6. South African Department of Environmental Affairs (DEA). 2012 South Africa environment outlook: Chapter 5: Air quality. Pretoria: DEA; 2012. http://soer.deat.gov.za/dm_documents/Chapter_5__Air_Quality_4kp8X.pdf
7. World Health Organization (WHO). Health effects of particulate matter: Policy implications for countries in the eastern Europe, Caucasus and central Asia. J Korean Med Assoc. 2013;50(2):1–20. http://www.euro.who.int/__data/assets/pdf_file/0006/189051/Health-effects-of-particulate-matter-final-Eng.pdf
8. World Health Organization (WHO): Occupational and Environmental Health Team. WHO Air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfur dioxide: Global update 2005: Summary of risk assessment. Geneva: WHO Press; 2006. https://apps.who.int/iris/handle/10665/69477
9. Chameides WL, Fehsenfeld F, Rodgers MO, Cardelino C, Martinez J, Parrish D, et al. Ozone precursor relationships in the ambient atmosphere. J Geophys Res Atmos. 1992;97(D5):6037–6055. https://doi.org/10.1029/91JD03014
10. Intergovernmental Panel on Climate Change (IPCC). Climate change 2001: The scientific basis. New York: Cambridge University; 2001. https://www.ipcc.ch/report/ar3/wg1/
11. Josipovic M. Acidic deposition emanating from the South African Highveld: A critical levels and critical loads assessment [PhD thesis]. Johannesburg: University of Johannesburg; 2009.
12. Georgoulias AK, Van der A RJ, Stammes, P, Boersma, KF, Eskes HJ. Trends and trend reversal detection in 2 decades of tropospheric NO_2 satellite observations. Atmos Chem Phys. 2019;19:6269–6294. https://doi.org/10.5194/acp-19-6269-2019
13. Lourens A, Butler T, Beukes J, Van Zyl P, Beirle S, Wagner T, et al. Re-evaluating the NO_2 hotspot over the South African Highveld. S Afr J Sci. 2012;108(11/12):1–6. https://doi.org/10.4102/sajs.v108i11/12.1146
14. Health Effects Institute (HEI). State of global air 2018 [webpage on the Internet]. No date [cited 2018 Nov 15]. Available from: https://www.stateofglobalair.org/
15. Lin J-T. Satellite constraint for emissions of nitrogen oxides from anthropogenic, lightning and soil sources over East China on a high-resolution grid. Atmos Chem Phys. 2012;12(6):2881–2898. https://doi.org/10.5194/acp-12-2881-2012
16. Pickering K, Bucsela E, Allen D, Ring A, Holzworth RA, Krotkov N. Estimating of lightning NO_x production based on OMI NO_2 observations over the Gulf of Mexico. J Geophys Res Atmos. 2014;121:8668–8691. https://doi.org/10.1002/2015JD024179
17. Beirle S, Platt U, Wagner TK. Potential of monitoring nitrogen oxides with satellite instruments. Proceedings of the 2006 EUMETSAT Meteorological Satellite Conference; 2006 June 12–16; Helsinki, Finland. Available from: https://www.eumetsat.int/website/home/News/ConferencesandEvents/PreviousEvents/DAT_2043223.html
18. Kang D, Pickering K. Lightning NO_x emissions and the implications for surface air quality over the contiguous United States. J Air Waste Manag Assoc. 2018:1–6.
19. Labrador LJ, Kuhlmann R, Lawrence MG. The effects of lightning-produced NO_x and its vertical distribution on atmospheric chemistry: Sensitivity simulations with MATCH-MPIC. Atmos Chem Phys. 2005;5(7):1815–1834. https://doi.org/10.5194/acp-5-1815-2005
20. Steiner A, Tonse S, Cohen R, Goldstein A, Harley R. Influence of future climate and emissions on regional air quality in California. J Geophys Res. 2006;111(18):1–22. https://doi.org/10.1029/2005JD006935
21. Tagaris E, Manomaiphiboon K, Liao K-J, Leung L, Woo J-H, He S, et al. Impacts of global climate change and emissions on regional ozone and fine particulate matter concentrations over the United States. J Geophys Res. 2007;112(14):1–11. https://doi.org/10.1029/2006JD008262
22. Banerjee A, Archibald A, Maycock A, Telford P, Abraham N, Yang X, et al. Lightning NO_x a key chemistry-climate interaction: Impacts of future climate change and consequences for tropospheric oxidising capacity. Atmos Chem Phys. 2014;14(18):9871–9881. https://doi.org/10.5194/acp-14-9871-2014
23. Finney DL, Doherty RM, Wild O, Young PJ, Butler A. Response of lightning NO_x emissions and ozone production to climate change: Insights from the Atmospheric Chemistry and Climate Model Intercomparison Project. Geophys Res Lett. 2016;43(10):5492–5500. https://doi.org/10.1002/2016GL068825
24. DeCaria AJ, Pickering KE, Stenchikov GL, Ott LE. Lightning-generated NO_x and its impact on tropospheric ozone production: A three-dimensional modelling study of a Stratosphere-Troposphere Experiment: Radiation, Aerosols and Ozone (STERAO-A) thunderstorm. J Geophys Res. 2005;110(14):1–13. https://doi.org/10.1029/2004JD005556
25. Price C, Penner J, Prather M. NO_x from lightning: 1. Global distribution based on lightning physics. J Geophys Res. 1997;102(5):5929–5941. https://doi.org/10.1029/96JD03504
26. Lawrence MG, Chameides WL, Kasibhatla PS, Levy II H, Moxim W. Lightning and atmospheric chemistry: The rate of atmospheric NO production. Boca Raton, FL: CRC Press; 1995.
27. Huntrieser H, Feigl C, Schlager H, Schroder F, Gerbig C, Van Velthoven P, et al. Airborne measurements of NO_x, tracer species, and small particles during the European Lightning Nitrogen Oxide Experiment. J Geophys Res Atmos. 2002;107(D11):ACH5-1–ACH5-24. https://doi.org/10.1029/2000JD000209
28. Wang Y, DeSilva AW, Goldenbaum GC, Dickerson RR. Nitric oxide production by simulated lightning: Dependence on current, energy, and pressure. J Geophys Res. 1998;103(D15):19149–19159. https://doi.org/10.1029/98JD01356
29. Holmes C, Brook M, Krehbiel P, McRory R. On the power spectrum and mechanism of thunder. J Geophys Res. 1971;76:2106–2115. https://doi.org/10.1029/JC076i009p02106

30. Gollardo L, Cooray V. Could cloud-to-cloud discharges be as effective as cloud-to-ground discharges in producing NO_x? Tellus. 1996;48B:641–651. https://doi.org/10.1034/j.1600-0889.1996.t01-4-00003.x

31. Cooray GV. Energy dissipation in lightning flashes. J Geophys Res. 1997;102:21401–21410. https://doi.org/10.1029/96JD01917

32. DeCaria AJ, Pickering KE, Stenchikov GL, Scala JR, Stith JL, Dye JE, et al. A cloud-scale model study of lightning-generated NO_x in an individual thunderstorm during STERAO-A. J Geophys Res. 2000;105(D9):11601–11616. https://doi.org/10.1029/2000JD900033

33. Ott LE, Pickering KE, Stenchikov GL, Allen DJ, DeCaria AJ, Ridley B, et al. Production of lightning NO_x and its vertical distribution calculated from three-dimensional cloud-scale chemical transport model simulations. J Geophys Res. 2010;115(4):1–19. https://doi.org/10.1029/2009JD011880

34. Biazar A, McNider R. Regional estimates of lightning production of nitrogen oxides. J Geophys Res. 1995;100(11):22861–22874. https://doi.org/10.1029/95JD01735

35. Beirle S, Huntrieser H, Wagner T. Direct satellite observation of lightning-produced NO_x. Atmos Chem Phys. 2010;10(22):10965–10986. https://doi.org/10.5194/acp-10-10965-2010

36. Ojelede M, Annegarn H, Price C, Kneen M, Goyns P. Lightning-produced NO_x budget over the Highveld region of South Africa. Atmos Environ. 2008;42(22):5706–5714. https://doi.org/10.1016/j.atmosenv.2007.12.072

37. Gijben M. The lightning climatology of South Africa. S Afr J Sci. 2012;108(3/4), Art. #740. https://doi.org/10.4102/sajs.v108i3/4.740

38. Vaisala. CP series CP7000 CP8000 user's guide. Helsinki: Vaisala Oyj; 2004.

39. Gill T. Initial steps in the development of a comprehensive lightning climatology of South Africa [MSc thesis]. Johannesburg: University of the Witwatersrand; 2008.

40. Schumann U, Huntrieser H. The global lightning-induced nitrogen oxides source. Atmos Chem Phys. 2007;7:3823–3907. https://doi.org/10.5194/acpd-7-2623-2007

41. Boccippio D, Cummins K, Christian H, Goodman S. Combined satellite and surface-based estimation of the Intracloud-cloud-to-ground lightning ratio over the continental United States. Mon Weather Rev. 2001;129:108–122. https://doi.org/10.1175/1520-0493(2001)129%3C0108:CSASBE%3E2.0.CO;2

42. Prentice SA, Mackerras D. The ratio of cloud to cloud-ground lightning flashes in thunderstorms. J Appl Meteorol Climatol. 1977;16(5):545–550. https://doi.org/10.1175/1520-0450(1977)016%3C0545:TROCTC%3E2.0.CO;2

43. Mackerras D, Darveniza M, Orville RE, Williams ER, Goodman SJ. Global lightning: Total, cloud and ground flash estimates. J Geophys Res. 1998;103:19791–19809.

44. Nesbitt SW, Zhang R, Orville R. Seasonal and global NO_x production by lightning estimated from the Optical Transient Detector (OTD). Tellus B: Chem Phys Meteorol. 2000;52(5):1206–1215. http://dx.doi.org/10.1034/j.1600-0889.2000.01121.x

45. Dye JE, Ridley BA, Skamarock W, Barth M, Venticinque M, Defer E, Blanchet P, Thery C, Laroche P, Baumann K, Hubler G, Parrish DD, Ryerson T, Trainer M, Frost G, Holloway JS, Matejka T, Bartels D, Fehsenfeld FC, Tuck A, Rutledge SA, Lang T, Stith J, Zerr R. An overview of the Stratospheric-Tropospheric Experiment: Radiation, Aerosols, and Ozone (STERAO)-Deep Convection experiment with results for the July 10, 1996 storm. J Geophys Res. 2000;105:10023–10045. https://doi.org/10.1029/1999JD901116

46. Bond DW, Zhang R, Tie X, Brasseur G, Huffines G, Orville RE, Boccippio DJ. NOx production by lightning over the continental United States. J Geophys Res. 2001;106(D21): 27701–27710. https://doi.org/10.1029/2000JD000191

47. Cummings KA, Huntemann TL, Pickering KE, Barth MC, Skamarock WC, Höller H, et al. Cloud-resolving chemistry simulation of a Hector thunderstorm. Atmos Chem Phys. 2013;13(5):2757–2777. https://doi.org/10.5194/acp-13-2757-2013

48. Sarkar D. Lattice: Multivariate data visualization with R. New York: Springer; 2008. https://doi.org/10.1007/978-0-387-75969-2

49. Beirle S, Spichtinger N, Stohl A, Cummins K, Turner T, Boccippio D, Cooper OR, Wenig M, Grzegorski M, Platt U, Wagner T. Estimating the NOx produced by lightning from GOME and NLDN data: A case study in the Gulf of Mexico. Atmos Chem Phys Discuss. 2005;5(6):11295–11329. https://hal.archives-ouvertes.fr/hal-00301908

50. Fehr T, Holler H, Huntrieser H. Model study on production and transport of lightning-produced NO_x in a EULINOX supercell storm. J Geophys Res. 2004;109:D09102. https://doi.org/10.1029/2003JD003935

51. Beirle S, Spichtinger N, Stohl A, Cummins KL, Turner T, Boccippio D, Cooper OR, Wenig M, Grzegorski M, Platt U, Wagner T. Estimating the NO_x produced by lightning from GOME and NLDN data: A case study in the Gulf of Mexico. Atmos Chem Phys. 2006;6(4):1075–1089. https://doi.org/10.5194/acp-6-1075-2006

52. Bucsela EJ, Pickering KE, Huntemann TL, Cohen RC, Perring A, Gleason JF, Blakeslee RJ, Albrecht RI, Holzworth R, Cipriani JP, Vargas-Navarro D, Mora-Segura I, Pacheco-Hrnandez A, Laporte-Molina S. Lightning-generated NOx seen by the Ozone Monitoring Instrument during NASA's Tropical Composition, Cloud and Climate Coupling Experiment (TC4). J Geophys Res. 2010;115:D00J10. https://doi.org/10.1029/2009JD013118

53. Bhavika B. The influence of terrain elevation on lightning density in South Africa [MSc thesis]. Johannesburg: University of Johannesburg; 2007.

54. Kotroni V, Lagouvardos K. Lightning occurrence in relation with elevation, terrain slope, and vegetation cover in the Mediterranean. J Geophys Res. 2008;113(21):1–7. https://doi.org/10.1029/2008JD010605

55. Bourscheidt V, Pinto O, Naccarato KP, Pinto IRCA. The influence of topography on the cloud-to-ground lightning density in South Brazil. Atmos Res. 2009;91(2–4):508–513. https://doi.org/10.1016/j.atmosres.2008.06.010

56. Vogt BJ, Hodanish SJ. A high-resolution lightning map of the state of Colorado. Mon Weather Rev. 2014;142:2353–2360. https://doi.org/10.1175/MWR-D-13-00334.1

57. Freiman MT, Piketh SJ. Air transport into and out of the industrial Highveld region of South Africa. J Appl Meteorol. 2003;42(7):994–1002. https://doi.org/10.1175/1520-0450(2003)042%3C0994:ATIAOO%3E2.0.CO;2

58. South African Department of Environmental Affairs (DEA). The medium-term review of the 2011 Highveld Priority Area (HPA): Air quality management plan. Pretoria: DEA; 2015. https://lifeaftercoal.org.za/wp-content/uploads/2016/02/HPA-AQMP-Midterm-review-Draft-Report_February-2016.pdf

59. Center for Environmental Rights, Groundwork and Highveld Environmental Justice Network. Life after coal/Impilo ngaphandle kwamalahle. Broken promises: The failure of the Highveld Priority Area [https://lifeaftercoal.org.za]. c2017 [2017 Nov 30]. Available from: https://lifeaftercoal.org.za/virtual-library/publications/broken-promises-the-failure-of-the-highveld-priority-area

60. Collier A, Hughes A, Lichtenberger J, Steinbach P. Seasonal and diurnal variation of lightning activity over southern Africa and correlation with European whistler observations. Ann Geophys. 2006;24(2):529–542. https://doi.org/10.5194/angeo-24-529-2006

61. Janssens-Maenhout G, Crippa M, Guizzardi D, Muntean M, Schaaf E. Emissions database for global atmospheric research version v4.2 (grid-maps). c2011 [cited 2017 Mar 12]. Available from: http://data.europa.eu/89h/jrc-edgar-jrc-edgarv42_gridmaps

62. South African Department of Environmental Affairs (DEA). GHG inventory for South Africa 2000 – 2010. Pretoria: DEA; 2014. https://unfccc.int/resource/docs/natc/zafnir1.pdf

63. Delmas R, Serça D, Jambert C. Global inventory of NOx sources. Nutr Cycling Agroecosyst. 1997;48(1–2):51–60. https://doi.org/10.1023/A:1009793806086

64. Bruwer A, Kornelius G. Modelling the effects of biogenic NOx emissions on the South African Highveld and Waterberg regions. Water Air Soil Pollut. 2017;228(9), Art. #326. https://doi.org/10.1007/s11270-017-3526-y

AUTHOR:
Linda Godfrey[1,2]

AFFILIATIONS:
[1]Waste RDI Roadmap, Council for Scientific and Industrial Research, Pretoria, South Africa
[2]Unit for Environmental Sciences and Management, North-West University, Potchefstroom, South Africa

CORRESPONDENCE TO:
Linda Godfrey

EMAIL:
LGodfrey@csir.co.za

DATES:
Received: 21 Sep. 2020
Revised: 13 Apr. 2021
Accepted: 13 Apr. 2021
Published: 29 Sep. 2021

HOW TO CITE:
Godfrey L. Quantifying economic activity in the informal recycling sector in South Africa. S Afr J Sci. 2021;117(9/10), Art. #8921. https://doi.org/10.17159/sajs.2021/8921

ARTICLE INCLUDES:
☒ Peer review
☐ Supplementary material

DATA AVAILABILITY:
☐ Open data set
☐ All data included
☒ On request from author(s)
☐ Not available
☐ Not applicable

EDITOR:
Jenni Case

KEYWORDS:
informal waste reclaimers, earnings, recycling, Extended Producer Responsibility

FUNDING:
None

Quantifying economic activity in the informal recycling sector in South Africa

Informal waste reclaimers are a key part of South Africa's recycling economy, being responsible for around 51% of all paper and packaging waste collected in South Africa in 2017. Active in the waste and recycling landscape for more than three decades, their activity predates the earliest voluntary paper and packaging Extended Producer Responsibility schemes. However, these voluntary schemes have been instrumental in scaling South Africa's recycling economy. Investment by brand owners, retailers, converters and recyclers has helped develop local end-use markets, creating a demand for paper and packaging recyclables and a resultant increase in their collection. An analysis of tonnage and price data shows that the mean estimate of money paid by the private sector to the informal waste sector through the purchase of recyclables at intermediaries such as buy-back centres, was ZAR625 million in 2012, increasing to ZAR872 million in 2017. This private sector 'investment' in the local recycling economy has led to direct and indirect job creation and improved livelihoods, particularly for a large, well-established and effective informal waste sector, and has indirectly funded municipal waste diversion strategies, saving municipalities in both the collection and disposal of waste.

Significance:

- Informal waste reclaimers make a significant contribution to the diversion of paper and packaging recyclables from landfill to recycling in South Africa.
- Limited information exists on the earnings of informal reclaimers.
- This paper provides new insights on the annual financial payments made by the private sector to informal reclaimers for the collection of recyclables.

Introduction

South Africa's 2nd National Waste Management Strategy[1] set the goal that 'all metropolitan municipalities, secondary cities and large towns have initiated separation at source programmes' by 2016, with '25% of recyclables diverted from landfill sites for re-use, recycling or recovery'. This goal was not achieved, with only a handful of cities and towns to date having implemented separation-at-source programmes, none of which cover the full spatial extent of the city or town. However, informal waste reclaimers, also referred to as waste pickers, have operated in this space for years, collecting recyclables through what Samson[2] calls 'separation-outside-source'.

Informal waste reclaimers have been active in the collection of paper and packaging recyclables in South Africa for more than three decades, with early policy documents such as the 1st National Waste Management Strategy[3] and the White Paper on Integrated Pollution and Waste Management[4] referencing their activity in the waste sector. Some of the earliest research on the informal waste sector in South Africa is thought to be that of De Kock[5] and Naidoo[6], who recognised waste picking as a means of survival for the 'unemployed, urban poor' who had no alternative means to financially support themselves and their families.

However, informal waste reclaimers play a key, although largely unrecognised, role in the South African recycling value chain.[2,7,8] They provide an important bridge between the municipal service chain and the formal, private sector value chain in developing countries, effectively moving 'waste' from kerbside and landfill into the recycling economy.[9,10] Early data published by the packaging sector, suggested that 80–90% (by weight) of post-consumer paper and packaging collected in South Africa in 2014, was done so by informal waste reclaimers at little to no cost to government or business.[10,11]

Recycling has been taking place in South Africa for many decades. Steelrec, the predecessor to Collect-a-Can which supports the collection and recycling of used beverage cans, was established in 1976 by Metal Box and Crown Cork (the predecessors of Nampak and ArcelorMittal).[12,13] Waste sorting facilities in major metropolitan areas such as Johannesburg and Pretoria were already operating in the 1970s.[14] However, as noted by Collect-a-Can[13], local end-use markets were limited in the 1970s, resulting in the stockpiling of beverage cans. Early paper and packaging (input) recycling rates were low, typically ≤20% in 1990, with the exception of paper which had already reached a recovery rate of 29%.[10]

For many waste streams, recycling is a negative value-added process requiring some form of market intervention, such as subsidies.[15-17] This is often due to the cost of collection, which accounts for a disproportionate share of the overall cost of recycling.[16] As noted by Nahman[18], the purpose of Extended Producer Responsibility (EPR) is, amongst others, to provide funding for the provision of incentives, subsidies and infrastructure to increase the supply of recyclables. As with recycling economies in many developed countries, the South African paper and packaging recycling economy has been subsidised by brand owners, converters and retailers through a number of voluntary EPR schemes, managed by Producer Responsibility Organisations (PROs)[19] such as PETCO, Polyco, FibreCircle, MetPac and TGRC. The earliest established PROs in South Africa include Collect-a-Can, established in

1993, and PETCO, which was established in 2004. Collect-a-Can was able to increase beverage can collection rates from 18% in 1993 to 72% in 2015.[20] Since the establishment of PETCO, post-consumer bottle PET recycling increased from 16% in 2005 to 55% in 2016.[21] Prior to 2000, only around 2% of PET was collected annually for recycling. According to the Sustainability Business Handbook[22]: 'At the time, PET recycling was a small-scale and uncoordinated activity that functioned without any industry direction or intervention.'

While the informal collection of recyclables predates the establishment of voluntary EPR schemes in South Africa, it appears that these EPR schemes, and the mandated PROs, were able to significantly scale-up recycling activities, by directing private sector funding into the development of local recycling infrastructure and local end-use markets. This created demand for recyclables, further fuelling collection and growing local recycling rates. As at 2017, an estimated 58.2% of paper and packaging put into the South African market was collected for recycling.[10,19] This investment in the local recycling industry is expected to continue as national government moves to implement mandatory EPR. The draft Paper and Packaging Industry Waste Management Plan submitted to government outlined a planned investment of ZAR2.85 billion over the period 2019–2023.[19]

The informal waste sector has, however, asserted that it too has been subsidising South Africa's recycling economy, with little compensation. As noted by Mr Kodisang[23] of WIEGO,

> *The invisible subsidy we don't see, is the subsidy of free labour of our informal waste reclaimers, a marginalized sector of our community forced to seek out a livelihood through the resource value only from the sale of recyclables. We can't have a recycling industry based on the free labour of poor people, this has to change.*

Informal reclaimers have, to date, typically only been paid the market-related, although highly variable, resource value upon sale of recyclables at intermediaries such as buy-back centres. Informal reclaimers are not paid for the collection service they provide, unlike many formal collectors. This issue of compensation for the currently free 'labour costs' of South Africa's informal waste sector, is addressed in the Waste Picker Integration Guideline for South Africa[24], and is currently the subject of a number of pilot projects. However, it is likely that payment for both the resource value and collection service will have significant cost implications for South Africa's recyclers, and ultimately producers.

A number of studies have explored the individual daily, weekly or monthly earnings of waste reclaimers in South Africa.[7,25,26,27] However, the author is unaware of research that has attempted to quantify the total annual payment made by the private sector to the informal waste sector. In this paper, therefore, an attempt is made to calculate the amount of money paid to the informal waste sector in South Africa, through intermediaries such as buy-back centres. Framed differently, the paper addresses the question: How much money has the informal waste sector earned through value-creation from paper and packaging recyclable waste in South Africa?

Methods, assumptions and data sources

To determine the payments made to the informal waste sector, tonnage and price data were sourced for South Africa. Because price data formed the limiting data set, the calculations were made for the years 2012 and 2017. Further information on each of the data sets is provided in the following sections.

Tonnages collected

Accurate, reliable waste data in South Africa, like most developing countries, are limited.[28,29] As such, data on the tonnages of paper and packaging recyclables collected in South Africa in 2012 and 2017 were sourced from industry[19,30] (Table 1).

The data provided annually by BMi Research is for high-level paper and packaging categories. As the price data were available for sub-categories, further data on waste tonnages were sourced from the individual PROs, through their annual reports (Table 2).

Plastic

High-level plastic data were sourced from BMi[30] and Packaging SA[31], with the disaggregated data by polymer type, sourced from Plastics SA[32,33]. For 2012, both packaging and non-packaging tonnages were used.[32] The tonnages of waste plastic exported, as reported by Plastics SA, while it should be included in the calculations (as it was collected in South Africa despite the end-destination for reprocessing), could not be included, as the export tonnages were not assigned to specific polymer groups. The excluded export tonnages account for 5.2% of the waste plastic collected in 2012.

The 2017 collected plastics data were more challenging to obtain. The Plastics SA report[33] notes 'Since 2015, plastics report on input figures, aligned with international reporting methods', which would suggest that all figures provided in the report are collection figures. However, the different figures provided in the report for (only) PET (67 872 recycled and 93 235 collected), suggests that the data in the report are in fact output (recycled) figures. As such, the BMi data for total plastic collected, total polyolefins collected and total PET collected were used.[31] The total polyolefins figure was subdivided into the sub-polymers using the ratios of recycled tonnages, assuming the same 'wastage' across the different polyolefins.

Paper

Detailed tonnages for the various sub-grades of paper were accessed from the Paper Recycling Association of South Africa.[34,35] It is noted that there are minor discrepancies in tonnage data between BMi, Packaging SA and PRASA. For this reason, the data were sourced directly from the PRO responsible for paper recycling in South Africa. Only the 'paper recovered in South Africa' data were used (excluding imports), as it is unlikely that the local informal waste sector would have been involved in the collection of imported paper.

Table 1: Paper and packaging consumed and collected in South Africa[19]

Packaging type	2012			2017		
	Consumption ('000 tonnes)	Collected ('000 tonnes)	% Collected	Consumption ('000 tonnes)	Collected ('000 tonnes)	% Collected
Plastic	734.1	291.0	39.6	867.8	395.1	45.5
Paper	2051.8	1151.3	56.1	2067.1	1393.6	67.4
Metal	230.2	148.7	64.6	183.3	138.9	75.8
Glass	865.4	339.2	39.2	758.8	330.7	43.6
Total	**3881.4**	**1930.2**	**49.7**	**3877.0**	**2258.3**	**58.2**

Metal

High level data for metals were also sourced from BMi. According to Packaging SA[19]: 'Since 2013, more than R1.258 billion in capital investments have been made by industry players … to convert the beverage can industry from steel to aluminium'. As such, it has been assumed that the metal collection data given for 2012[30], do not yet include aluminium. The tonnages for steel/tinplate and aluminium for 2017 were sourced from BMi.[31]

Glass

Only one data set was provided by BMi for 'glass'.[30,31] No further subdivision of glass by colour was used, as price data also were not available for sub-categories.

Prices for recyclables

Recyclable price data for the year 2012, as paid-for recyclables at buy-back centres across South Africa, were sourced from Viljoen[36] and Viljoen et al.[37] Price data for 2017 were sourced from Schenck et al.[38] The 2012 data were captured at buy-back centres in 11 of South Africa's large cities, i.e. Bloemfontein, Cape Town, Durban, East London, Johannesburg, Kimberley, Mafikeng, Mbombela, Pietermaritzburg, Gqeberha (Port Elizabeth) and Pretoria. The 2017 data were captured at the provincial level. The average provincial price per paper and packaging grade was used here, i.e. Gauteng, Eastern Cape, Western Cape, KwaZulu-Natal, Mpumalanga, Free State and Northern Cape.

Where no price data were available for a specific sub-category of paper or packaging, the lowest value from another city/town or region was applied. As such, the resultant total values are likely to be conservative underestimates of the money paid to informal reclaimers.

The 2017 price data contained much greater detail in prices for sub-categories compared to the 2012 data. However, the 2017 price detail for sub-categories could often not be used, as the matching tonnage data were not available, e.g. for 2017 prices, the average for the five different low-density plastic grades was used. Because there is no recorded data on the breakdown of the tonnages of clear/brown/green PET bottles collected in South Africa (Scholtz C 2020, written communication, 23 December) it is assumed that the majority of bottles collected (driven by the higher value) are clear PET bottles and as such, the prices for clear bottles have been used.

Informally collected recyclables

There is also very little reliable data on the number of informal waste reclaimers in South Africa, and the tonnages of recyclable waste collected by the informal sector. Available data published by BMi[30,39,40] on the percentage of paper and packaging recyclables collected by informal trade, including sub-categories where available, are shown in Table 2. 'Informal trade' is defined by BMi[40] as 'informal businesses that sell directly to you'.

Table 2: Percentage of paper and packaging recyclables collected through informal trade

Packaging type	2012	2014	2015
Plastic	68%	17.1%	4%
PET	90%	85%	85%
Paper	Mostly informal	Mostly informal	30%
Metal	30–40%	30–40%	–
Glass	80%	–	–

PET, polyethylene terephthalate

The reports also note that 'informal trade is responsible for the bulk of collection'[30,39]. None of the PROs were able to provide accurate figures on the percentage of paper and packaging collected by informal waste reclaimers.

According to Plastics SA[41], an estimated 70% in 2011 and 58% in 2012 of recyclable plastic waste was sourced from post-consumers and landfills. The Plastics SA report[33] shows that, in 2017, this number had increased, as 'The largest quantity of recyclables, 74%, was obtained from landfill and other post-consumer sources'. As this is where informal reclaimers work, it is assumed that a high percentage of this post-consumer waste is collected by the informal sector. The same percentage (68%) was therefore used for plastic (with the exception of PET) for both 2012 and 2017. The percentage of PET plastic collected by informal reclaimers is reported as being higher, due to its higher value (prices). For paper, the BMi reports[30,39] noted 'mostly informal' for paper collection (assumed here to be around 80%), which decreased to 30% in 2015. While paper and cardboard are highly sought after by informal reclaimers, the Paper Manufacturers Association of South Africa has suggested a maximum of 50% collection by the informal sector in 2017 for most paper grades, with corrugated paper as low as 33% (Molony J 2020, written communication, 10 January). These figures are based on what they know to be collected by the formal sector. Metals are reported as 30–40% collected by informal trade; however, this number may in fact be higher given the prices for metal, in particular aluminium, and the resultant demand by the informal sector. Using this figure of 30–40% will also likely result in an underestimate in the calculation of money paid to informal waste reclaimers.

Based on the information provided by the PROs, the contribution by the informal sector to the local recycling economy, in terms of the reported percentage of recyclable waste collected, appears to have declined over the years. It is not clear if this is due to improved data by the private sector, or whether this is a real decrease in the informal sector's contribution, due to erosion by the formal sector.

In terms of the number of informal reclaimers, the Waste Sector Survey[42] indicated that the informal waste sector could be 2–3 times larger than the formal waste sector, at 60 000–90 000 people. The South African Department of Environmental Affairs estimated the number at 62 147.[43] As there is no agreement on the number of informal reclaimers in South Africa, assessments have been made for two different scenarios – 60 000 and 90 000.

Uncertainty

The uncertainty in the percentage of recyclable waste collected by the informal sector has an impact on the results presented in the following section. This is particularly evident for waste paper, where a significant decline in the percentage of paper collected by informal reclaimers between 2012 (80%) and 2017 (50%) was reported by the sector. For this reason, a range in figures is provided in the following section.

Results and discussion

Based on the above assumptions and available data sets, the results (presented in Tables 3 and 4 and Figure 1) show that between ZAR383 million and ZAR882 million was paid to the informal waste sector, via buy-back centres, in 2012. This increased to between ZAR423 million and ZAR1.54 billion in 2017. The mean estimate of total annual payments made to the informal waste sector in South Africa, via buy-back centres, is estimated to have increased from ZAR625 million in 2012 to ZAR872 million in 2017. Without data on the geographic distribution of collected recyclables, it is not possible to say where in this income range the answer lies. However, given that a large percentage of the paper and packaging recyclers reside in the major metropolitan areas, 'in close proximity to their incoming material sources'[33], e.g. Gauteng, where prices are also generally higher than in other parts of the country, it is likely that these average values are conservative.

According to the data provided by industry, the total tonnages collected by the informal sector decreased from 1.45 Mt in 2012 to 1.10 Mt in 2017. While the overall percentage of paper and packaging collected by the informal sector in South Africa decreased from 76% in 2012 to 51% in 2017, the average money paid to informal reclaimers increased by around 40% over the same period, highlighting the increase in recyclable prices over this 5-year period. The waste streams which supported the largest average payments to informal waste reclaimers in 2017 were PET plastic, corrugated paper and aluminium (Table 4).

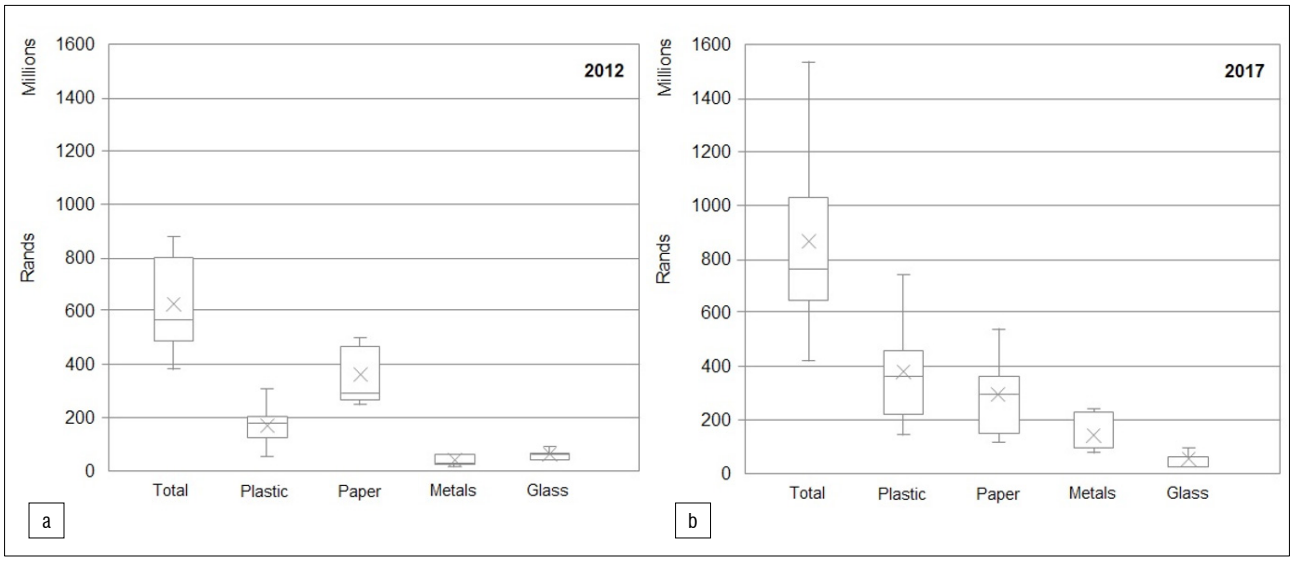


Figure 1: Estimated range (total and per recyclable stream) of money paid by business, via buy-back centres, to informal waste reclaimers in (a) 2012 and (b) 2017.

Table 3: Estimated payment to informal reclaimers in South Africa via buy-back centres in 2012

Recyclable waste type	Apparent South African consumption (ASAC) (t)	Total collected†		Range in prices paid to informal reclaimers at buy-back centres (ZAR/t)	Estimate of waste quantities collected by informal reclaimers		Estimate of total money paid by business (via buy-back centres) to informal waste reclaimers		
		Tonnes (t)	% ASAC		% Collected informally	Tonnes collected informally	Minimum (ZAR)	Maximum (ZAR)	Average (ZAR)
Plastic	**710 700**	**268 548**	**37.8**			**193 674**	**55 470 408**	**307 692 904**	**171 664 100**
PE-LD/LLD		98 971		330–1380	68%	67 300	22 209 092	92 874 386	56 471 053
PE-HD		45 949		330–1200	68%	31 245	10 310 956	37 494 384	20 905 960
PP		47 081		150–1630	68%	32 015	4 802 262	52 184 580	24 331 461
PET		50 280		330–2800	90%	45 252	14 933 160	126 705 600	55 618 822
PS		3 395		150–600	68%	2 309	346 290	1 385 160	440 733
PVC		16 812		150–5000	68%	11 432	1 714 824	57 160 800	11 951 804
Other		6 060		280–830	68%	4 121	1 153 824	3 420 264	1 944 268
Paper	**2 689 994**	**1 151 315**	**42.8**			**921 053**	**249 214 784**	**497 352 760**	**358 451 780**
Newspapers		116 831		170–330	80%	93 465	15 889 016	30 843 384	21 072 064
Magazines		66 574		180–350	80%	53 259	9 586 656	18 640 720	13 121 130
Corrugated		751 951		200–520	80%	601 561	120 312 160	312 811 616	219 296 255
Office, graphic papers		96 955		350–1850	80%	77 564	27 147 400	143 493 400	71 640 931
Mixed and other		119 005		130–1000	80%	95 204	12 376 520	95 204 000	33 321 400
Metal	**230 400**	**148 700**	**64.5**			**59 480**	**10 706 400**	**61 859 200**	**35 633 927**
Steel/tinplate	230 400	148 700		180–1040	40%	59 480	10 706 400	61 859 200	35 633 927
Aluminium		–		–	–	–	–	–	–
Glass	**865 500**	**339 200**	**39.2**	**150–330**	**80%**	**271 360**	**40 704 000**	**89 548 800**	**59 205 818**
Total	**4 496 594**	**1 907 764**	**42.4**	–	**76%**	**1 445 567**	**383 231 592**	**881 958 472**	**624 955 626**

PE, polyethylene; LD/LLD, low density/linear low density; HD, high density; PP, polypropylene; PET, polyethylene terephthalate; PS, polystyrene; PVC, polyvinyl chloride

†The tonnages collected and the collection as a percentage of consumption, as reported here, are lower than the official published figures, due to the focus here on what waste is likely to be collected by the informal sector (e.g. excludes imports where available).

Table 4: Estimated payment to informal reclaimers in South Africa via buy-back centres in 2017

Recyclable waste type	Apparent South African consumption (ASAC) (t)	Total collected†		Range in prices paid to informal reclaimers at buy-back centres (ZAR/t)	Estimate of waste quantities collected by informal reclaimers		Estimate of total money paid by business (via buy-back centres) to informal waste reclaimers		
		Tonnes (t)	% ASAC		% Collected informally	Tonnes collected informally	Minimum (ZAR)	Maximum (ZAR)	Average (ZAR)
Plastic	**867 800**	**395 077**	**45.5**			**284 496**	**144 116 771**	**742 841 126**	**379 118 608**
PE-LD/LLD		129 088		460–2000	68%	87 780	40 378 695	175 559 541	88 782 968
PE-HD		77 747		450–2880	68%	52 868	23 790 693	152 260 434	57 626 345
PP		58 112		600–3130	68%	39 516	23 709 679	123 685 491	41 887 099
PET		93 200		580–3570	85%	79 220	45 947 600	282 815 400	166 927 857
PS		6 609		–	68%	4 494	–	–	–
PVC		21 905		350–3880	68%	14 896	5 213 445	57 794 759	19 087 592
Other		8 415		680–1360	68%	5 722	3 891 176	7 782 352	4 806 747
Paper	**2 255 075**	**1 282 120**	**56.9**			**496 391**	**119 257 365**	**542 482 585**	**295 141 611**
Newspapers		146 509		100–640	50%	73 255	7 325 450	46 882 880	25 220 478
Magazines		30 955		300–930	50%	15 478	4 643 250	14 394 075	7 075 429
Corrugated		850 992		200–890	33%	280 827	56 165 472	249 936 350	150 844 411
Office, graphic papers		139 421		450–1900	50%	69 711	31 369 725	132 449 950	79 768 729
Mixed and other		114 242		100–1730	50%	57 121	5 712 100	98 819 330	32 232 564
Metal	**183 252**	**138 939**	**75.8**			**55 576**	**78 188 804**	**242 708 740**	**143 372 195**
Steel/tinplate	133 896	98 773		390–830	40%	39 509	15 408 588	32 792 636	19 867 483
Aluminium	49 356	40 166		3760–14 000	40%	16 066	60 409 664	224 929 600	123 504 712
Glass	**758 817**	**330 700**	**43.6**	**100–350**	**80%**	**264 560**	**26 456 000**	**92 596 000**	**54 045 829**
Total	**4 064 944**	**2 146 836**	**52.8**	–	**51%**	**1 101 023**	**423 137 396**	**1 537 636 907**	**871 678 242**

PE, polyethylene; LD/LLD, low density/linear low density; HD, high density; PP, polypropylene; PET, polyethylene terephthalate; PS, polystyrene; PVC, polyvinyl chloride

†The tonnages collected and the collection as a percentage of consumption, as reported here, are lower than the official published figures, due to the focus here on what waste is likely to be collected by the informal sector (e.g. excludes imports where available).

Table 5: Estimated average monthly income per informal waste reclaimer

Year	Mean annual total paid to informal waste reclaimers (ZAR)	Estimated annual informal reclaimer income (ZAR)		Equivalent monthly informal reclaimer income (ZAR)	
		60 000 reclaimers	90 000 reclaimers	60 000 reclaimers	90 000 reclaimers
2012	624 955 626	10 416	6 944	868	579
2017	871 678 242	14 528	9 685	1 211	807

Given the uncertainty in the informal waste sector data, and the large variations in the price data, the external validity[44] of this research was tested through triangulation with other studies. Because total country-wide payments made to informal reclaimers had not previously been calculated for South Africa, validity was tested against individual reclaimer earnings, thereby adopting a 'bottom-up' approach.

The estimated annual income per waste reclaimer in South Africa was calculated for two different scenarios (Table 5). At a conservative 60 000 informal waste reclaimers, the monthly income equates to ZAR868 in 2012, increasing to ZAR1211 in 2017. It is acknowledged that, while these are average amounts, the monthly income per reclaimer can vary significantly, depending on, for example, the number of hours worked; whether working at kerbside or on landfill; the geographic areas in which they work; and the type of material collected.[26,36] Based on the buy-back centre price data used in this study, it is clear that there are significant differences in the earnings of waste reclaimers, given variation in prices of different recyclables as well as the geographic area within which they work.

Blaauw et al.[45] note a range in the average monthly income of informal waste reclaimers working in Pretoria, South Africa (in 2010), based on a 'bad week' and a 'good week' of waste picking, of ZAR449 and ZAR1142 respectively. Viljoen et al.[27], who presented a national study of the socio-economic dynamics and vulnerability of street waste pickers (between April 2011 and June 2012), calculated the mean informal reclaimer income 'usually earned for a day's waste' at ZAR72.11. While the number of days worked by informal reclaimers varies significantly[26], a daily earning of ZAR72.11 could equate to a monthly income of around ZAR1440–R1730. Finally, a national study conducted by the South African Department of Environmental Affairs[46] showed a mean monthly earning of informal waste reclaimers of ZAR1430. These studies have noted the long work days of informal reclaimers, often 8 h or more per day, and the long work weeks, often 5–7 days per week.[26,27,46]

While there is considerable variation in the income earned by waste reclaimers, a comparison with these studies[26,27,45,46] suggests that this bottom-up triangulation approach provides findings of the same order of magnitude as the results from this study (Table 5). As comparison, the South African minimum wage[47] of ZAR20 per hour, based on an 8-h working day, five days per week, would generate a minimum monthly income of ZAR3500 per month.

The apparent decrease in the tonnages of recyclables informally collected and the monthly income paid to reclaimers were also discussed with representatives of the African Reclaimers Organisation, to assess whether these modelled results were actually 'observed' on the streets of South Africa. The main findings were supported by the African Reclaimers Organisation, highlighting the growing risk to informal waste reclaimers. The erosion in the informal sector's contribution to South Africa's recycling economy was noted by the African Reclaimers Organisation as being due to greater involvement by formal actors, which are often contracted by local municipalities. This risk is further compounded by an increase in the number of informal reclaimers, due to an escalating unemployment rate, which creates greater competition for easily accessible material. Reclaimers noted that due to this 'formal sector erosion', they were forced to diversify into other materials, including different types of plastic. The increasing contribution of plastic to the recyclables collected by the informal sector is evident in the results (Tables 3 and 4). Plastic made up 13.4% of the paper and packaging recyclables collected by the informal sector in 2012, which increased to 25.8% in 2017. While the tonnages of paper and packaging recyclables collected by informal reclaimers appears to have decreased between 2012 and 2017, this may not have been immediately noticeable to reclaimers, due to the increase in prices paid by buy-back centres over the same period, supporting increased earnings. However, with the slowing in the local recycling economy in 2019, the impact of the COVID-19 pandemic in 2020 and the resultant drop in local prices, this 'price buffer' may be eroded, with significant implications for the livelihoods of informal waste reclaimers.

Conclusions

In conclusion, using two data sets, the tonnages of paper and packaging collected for recycling, and the prices of various sub-categories of paper and packaging paid at buy-back centres around the country, estimates were determined of the total amount of money paid to the informal waste sector in 2012 and 2017. The mean estimate of total money paid by the private sector to the informal waste sector, through the sale of recyclables at intermediaries such as buy-back centres and scrap dealers in South Africa, was ZAR625 million in 2012. This increased to ZAR872 million in 2017. These actions have resulted in the collection (formal and informal) of over 2 million tonnes of paper and packaging waste, and ultimately the diversion of this waste away from municipal landfills, and from the environment. This private sector expenditure, combined with the work of the informal waste sector, not only saves municipalities in landfill airspace (estimated at between ZAR309 million and ZAR749 million in 2014)[11], but could also be considered an investment in municipal diversion strategies. However, if industry is to understand and harness the contribution of the informal and formal sectors to the South African recycling economy, and to better integrate informal reclaimers into the new mandatory EPR schemes, there is an urgent need for improved data collection.

Competing interests

I have no competing interests to declare.

References

1. South African Department of Environmental Affairs (DEA). National Waste Management Strategy [document on the Internet]. c2011 [cited 2019 Nov 14]. Available from: https://www.environment.gov.za/sites/default/files/docs/nationalwaste_management_strategy.pdf
2. Samson M. Whose frontier is it anyway? Reclaimer 'integration' and the battle over Johannesburg's waste-based commodity frontier. Capital Nat Social. 2020;31(4):60–75. https://doi.org/10.1080/10455752.2019.1700538
3. South African Department of Environmental Affairs and Tourism (DEAT). National Waste Management Strategy Version D. Pretoria: DEAT; 1999.
4. South African Department of Environmental Affairs and Tourism (DEAT). White Paper on Integrated Pollution and Waste Management for South Africa. A policy on pollution prevention, waste minimisation, impact management and remediation. Government Gazette No. 20978, Notice No. 227; 2000. [cited 2021 Mar 22]. Available from: https://www.environment.gov.za/sites/default/files/legislations/integrated_pollutionand_wastemanagement_0.pdf
5. De Kock RE. Garbage picking as a strategy for survival: A case study of a subsector of the informal sector [master's dissertation]. Durban: University of Natal; 1986. https://ukzn-dspace.ukzn.ac.za/bitstream/handle/10413/6327/De%20Kock_Rachelle_1987.pdf?sequence=1&isAllowed=y
6. Naidoo JSG. A study of the garbage pickers within the Port Shepstone disposal site [honours dissertation]. Durban: University of Zululand; 1994. http://uzspace.unizulu.ac.za/handle/10530/676
7. Schenck CJ, Blaauw PF, Swart EC, Viljoen JMM, Mudavanhu N. The management of South Africa's landfills and waste pickers on them: Impacting lives and livelihoods. Dev South Afr. 2019;36(1):80–98. https://doi.org/10.1080/0376835X.2018.1483822
8. Viljoen JMM. Economic and social aspects of street waste pickers in South Africa [doctoral dissertation]. Johannesburg: University of Johannesburg; 2014. http://hdl.handle.net/10210/12273
9. ECD. Working Party on Resource Productivity and Waste. In: Extended Producer Responsibility – Updated guidance ENV/EPOC/WPRPW(2015)16/FINAL. Paris: OECD; 2016.
10. Godfrey L, Oelofse S. Historical review of waste management and recycling in South Africa. Resources. 2018;6(4):57. https://doi.org/10.3390/resources6040057
11. Godfrey L, Strydom W, Phukubye R. Integrating the informal sector into the South African waste and recycling economy in the context of Extended Producer Responsibility. CSIR Briefing Note [document on the Internet]. c2016 [cited 2021 Mar 22]. Available from: https://wasteroadmap.co.za/wp-content/uploads/2020/03/informal_sector_2016.pdf
12. Nampak Limited. Annual report 2002 [webpage on the Internet]. c2002 [cited 2021 Mar 22]. Available from: http://www.nampak.com/Investors/Financial-Information
13. ArcelorMittal. A walk through history with Collect-a-Can [document on the Internet]. c2011 [cited 2021 Mar 22]. Available from: https://www.arcelormittalsa.com/Portals/0/A%20%20walk%20through%20history%20with%20Collect-a-Can.pdf
14. Noble RG. Solid waste research in South Africa. South African National Scientific Programmes report no. 4. Pretoria: Council for Scientific and Industrial Research; 1976. Available from: http://researchspace.csir.co.za/dspace/bitstream/handle/10204/2236/SANSP%20004%20.pdf?sequence=1&isAllowed=y
15. Matheson T. Disposal is not free: Fiscal instruments to internalize the environmental costs of solid waste. IMF Working Paper WP/19/283. Washington DC: International Monetary Fund; 2019.
16. Organisation for Economic Cooperation and Development (OECD). Extended Producer Responsibility: Updated guidance for efficient waste management. Paris: OECD Publishing; 2016. https://doi.org/10.1787/9789264256385-en
17. Pouikli K. Concretising the role of extended producer responsibility in European Union waste law and policy through the lens of the circular economy. ERA Forum. 2020;20:491–508. https://doi.org/10.1007/s12027-020-00596-9

18. Nahman A. Extended producer responsibility for packaging waste in South Africa: Current approaches and lessons learned. Resour Conserv Recycl. 2010;54(3):155–162. https://doi.org/10.1016/j.resconrec.2009.07.006

19. Packaging SA. Industry waste management plan: Federation of plans – Draft for public comment. Presented for: Packaging SA stakeholder engagement; 2018 July 20 – 2018 August 20.

20. Collect-a-Can. South Africa: Beverage can recovery rate 1993 to 2015 [webpage on the Internet]. No date [cited 2017 July 29]. Available from: http://www.collectacan.co.za/index.php/can-recycling/recovery-rate

21. PETCO. Future focus: A review of PETCO activities [webpage on the Internet]. c2016 [cited 2021 Mar 22]. Available from: https://petco.co.za/wp-content/uploads/2017/05/PETCO-2016-Annual-Review_web-version.pdf

22. Trialogue. The sustainable business handbook: Smart strategies for responsible companies. 4th ed. PETCO: Recycling PET plastic for a cleaner South Africa. Cape Town: Trialogue; 2007.

23. Kodisang E. Gauteng Waste to Wealth Seminar; 2019 June 07; Johannesburg, South Africa.

24. South African Department of Environment, Forestry and Fisheries (DEFF), Department of Science and Innovation (DSI). Waste picker integration guideline for South Africa. Pretoria: DEFF and DSI; 2020. Available from: https://wasteroadmap.co.za/wp-content/uploads/2021/02/Waste-Picker-Integration-Guidelines.pdf

25. Muller M. When necessity begets ingenuity: A study of informal waste recycling at Stellenbosch and Bellville, Cape Town [master's dissertation]. Cape Town: University of the Western Cape; 2015. Available from: http://etd.uwc.ac.za/xmlui/handle/11394/4945

26. Yu D, Blaauw D, Schenck R. Waste pickers in informal self-employment: Over-worked and on the breadline. Dev South Afr. 2020;37(6):971–996. https://doi.org/10.1080/0376835X.2020.1770578

27. Viljoen K, Blaauw D, Schenck R. Sometimes you don't make enough to buy food: An analysis of South African street waste pickers' income'. J Econ Financ Sci. 2018;11(1):a186. http://dx.doi.org/10.4102/jef.v11i1.186

28. Godfrey L, Nahman A. Are developing countries ready for first world policy instruments? In: Proceedings of Sardinia Eleventh International Waste Management and Landfill Symposium; 2007 October 1–5; Cagliari, Italy. Available from: https://researchspace.csir.co.za/dspace/bitstream/handle/10204/846/Godfrey1_2007.pdf?sequence=3&isAllowed=y

29. UN Environment Programme. Global waste management outlook [webpage on the Internet]. c2015 [cited 2019 Dec 12]. Available from: https://www.unenvironment.org/resources/report/global-waste-management-outlook

30. BMi Research. Recycling assessment report 2012. Unpublished report 2013 [courtesy of Packaging SA].

31. Packaging SA. Packaging market size and volumes collected, as compiled by BMI. Personal communication, Shabeer Jhetam, Packaging SA, May 2019.

32. Plastics SA. Plastics recycling survey 2012. Unpublished report 2013 [courtesy of Plastics SA].

33. Plastics SA. National plastics recycling survey 2017. Unpublished report 2018 [courtesy of Plastics SA].

34. Paper Recycling Association of South Africa (PRASA). South African paper recycling statistics for 2012 [document on the Internet]. c2013 [cited 2019 Dec 27]. Available from: https://recyclepaper.co.za/wp-content/uploads/2019/01/LB344-PRASA-Statistics-2012.pdf

35. Paper Recycling Association of South Africa (PRASA). Paper recycling in South Africa: Facts and statistics. Paper recycling statistics for 2017 [document on the Internet]. c2018 [cited 2019 Dec 27]. Available from: https://recyclepaper.co.za/wp-content/uploads/2019/01/Paper-recycling-statistics-2017-June-2018.pdf

36. Viljoen JMM. Economic and social aspects of street waste pickers in South Africa [doctoral dissertation]. Johannesburg: University of Johannesburg; 2014.

37. Viljoen J, Blaauw D, Schenck C. The opportunities and value-adding activities of buy-back centres in South Africa's recycling industry: A value chain analysis. Local Econ. 2019;34(3):294–315.

38. Schenck CJ, Barnes K, Viljoen JMM, Blaauw PF. Buy back research data sheet. Cape Town: University of the Western Cape; 2019.

39. BMi Research. Recycling assessment report 2014. Unpublished report 2015 [courtesy of Packaging SA].

40. BMi Research. Packaging South Africa report 2015. Unpublished report 2016 [courtesy of Packaging SA].

41. Plastics SA. Annual report 2012 [webpage on the Internet]. c2012 [cited 2019 Dec 23]. Available from: https://www.plasticsinfo.co.za/media-room/annual-reports/

42. South African Department of Science and Technology (DST). South African Waste Sector – 2012. An analysis of the formal private and public waste sector in South Africa. A National Waste RDI Roadmap for South Africa: Phase 1 Status Quo Assessment. Pretoria: DST; 2013.

43. South African Department of Environmental Affairs (DEA). Report on the determination of the extent and role of waste picking in South Africa [document on the Internet]. c2016 [cited 2019 Dec 23]. Available from: http://sawic.environment.gov.za/documents/5413.pdf

44. Leedy PD, Ormrod JE. Practical research: Planning and design. 8th ed. Hoboken, NJ: Pearson Prentice Hall; 2005.

45. Blaauw D, Pretorius A, Schenck R. The economics of urban waste picking in Pretoria. Afr Rev Econ Financ. 2019;11(2):129–164. Available from: https://african-review.com/view-paper.php?serial=20200617191809-257880

46. South African Department of Environmental Affairs (DEA). Report on the determination of the extent and role of waste picking in South Africa. Pretoria: DEA; 2016. Available from: http://sawic.environment.gov.za/documents/5413.pdf

47. National Minimum Wage Act, Act No. 9 of 2018, Republic of South Africa. Available from: https://www.gov.za/sites/default/files/gcis_document/201811/42060gon1303act9of2018.pdf

AUTHORS:
François B. van Schalkwyk[1]
Milandré H. van Lill[1]
Nico Cloete[1]

AFFILIATION:
[1]DSI-NRF Centre of Excellence in Scientometrics and STI Policy, Centre for Research on Evaluation, Science and Technology, Stellenbosch University, Stellenbosch, South Africa

CORRESPONDENCE TO:
François van Schalkwyk

EMAIL:
fbvschalkwyk@sun.ac.za

DATES:
Received: 06 Apr. 2021
Revised: 28 May 2021
Accepted: 10 June 2021
Published: 29 Sep. 2021

HOW TO CITE:
Van Schalkwyk FB, Van Lill MH, Cloete N. Brain circuity: The case of South Africa as a hub for doctoral education. S Afr J Sci. 2021;117(9/10), Art. #10674. https://doi.org/10.17159/sajs.2021/10674

ARTICLE INCLUDES:
☒ Peer review
☐ Supplementary material

DATA AVAILABILITY:
☐ Open data set
☐ All data included
☒ On request from author(s)
☐ Not available
☐ Not applicable

EDITOR:
Chrissie Boughey

KEYWORDS:
mobility, migration, doctoral studies, higher education, Africa

FUNDING:
National Research Foundation of South Africa (grant no. 91488)

Brain circuity: The case of South Africa as a hub for doctoral education

The production *and reproduction* of knowledge are important components of national development. As student mobility increases, globally and within Africa, so does the national diversity of students as they seek to further their postgraduate studies at the limited number of research universities in Africa. Knowledge migration is inevitably a relationship between nation states because migration is driven by push factors (such as the socio-economic conditions and opportunities) in the country of origin as well as by pull factors (such as the rules and incentives for entry, participation in postgraduate education and post-study residency), which are prerogatives of the host nation. In other words, migration and development must be understood in comparative terms. The brain drain perspective on migration and development takes mainly the perspective of the origin country into consideration. Migration and the loss of high-level skills are seen as detrimental to the development prospects of the country of origin. The brain circulation perspective moves the discussion forward by suggesting that there are residual returns to the country of origin. However, relatively little attention has been given to the impact of knowledge migrants on the host nation when the host is facing its own post-colonial development challenges. This is the dilemma facing South Africa as a hub for doctoral students from the rest of Africa: attracting top doctoral students from the rest of the continent to contribute to the country's knowledge capacity but at the expense of developing local talent, thereby setting up a complex tension between underdevelopment and development. Here we establish whether South Africa is maintaining its position as a PhD hub on the African continent and explore the extent to which the brain circulation argument holds up in the African context. We suggest that, given the current policy environment in South Africa, brain circuity is a more likely outcome, where brain circuity describes the flow of knowledge characterised by indirection and undesirable intricacy.

Significance:

- In the case of South Africa as a destination for doctoral students from the rest of Africa, neither the brain drain nor the brain circulation theories of student mobility hold; rather, students are confronted with an overly complex set of conditions resulting in brain circuity.
- Despite the continued attractiveness of South Africa as a destination for doctoral students, the tension between equity and development remains unresolved at the policy level, potentially undermining the circulation of knowledge for the benefit of all African countries.

Introduction

In both developed and developing countries, the importance of doctoral education has been growing.[1-3] This rise in importance can be attributed in part to the need for the academy to reproduce itself for its own survival. Doctoral graduates seed the academy. In addition to the survival of the academy, and more pertinent from a broader policy perspective, doctorates are seen to play an important developmental role. By ensuring the continued production and transfer of new knowledge, doctoral graduates contribute to socio-economic development. In sum, doctorates hold the key to both the reproduction and production of knowledge, which are important contributors to national development.[4]

The importance ascribed to doctoral studies needs to be viewed against the backdrop of at least two interlinked developments in higher education: diversity and mobility. Recent developments across the globe have reignited calls for more diverse institutions, including universities. At the same time, the mobility of students between countries, in part to support diversity aspirations, has been made possible by national and institutional policies, funding and incentives, as well as the ease of accessing information about study opportunities and of movement.

Both the mobility and diversity of university student populations most often fall under the rubric of the internationalisation of higher education.[5,6] For obvious reasons, prior to 1994, but also in early democratic South Africa, the internationalisation of universities in South Africa did not feature prominently in higher education policy.[7] As attention turned towards globalisation in post-apartheid South Africa, the internationalisation of postgraduate enrolments began to appear in various policy documents.[8] The National Plan for Higher Education[9], for instance, recommended that institutions increase recruitment of students from the Southern African Development Community (SADC), especially at the postgraduate level. South Africa's National Development Plan[10] envisaged South Africa as being capable of attracting a significant share of the international student population. And the most recent White Paper for Post-School Education and Training[11] noted that hosting large numbers of international students, especially SADC students, represents a major contribution by South Africa to the development of the sub-continent.

The South African National Development Plan[10] made several bold proposals, including for South African universities to graduate more than 100 doctoral candidates per one million of the population by 2030. This equates to an

increase from 1420 per annum in 2010 to 5250 doctorates per annum in 2030. At last count, in 2018, South Africa graduated 3344 doctoral candidates or 58 doctorates per million of the population.[12]

The National Development Plan also envisaged South Africa as a regional hub for higher education and training, particularly for students from the SADC. If South African higher education wants to achieve its doctoral graduate targets, it will have to enrol more students – not only from South Africa but also from the rest of Africa and the rest of the world. Sehoole[13], drawing on 2009 data, showed that South Africa was already establishing itself as a regional hub for doctoral candidates by attracting more doctoral students from Africa than from anywhere else globally. In 2015, Cloete et al.[14] also explored South Africa's status as a hub for doctoral students from the rest of Africa. They found that South Africa, despite a lack of a supportive national policy environment, was well placed to maintain and strengthen its position as the prime regional destination for doctoral candidates from the rest of Africa.[14]

In 2017, the government published the Policy Framework for Internationalisation of Higher Education in South Africa. The policy is a classic example of politically correct 'policy speak' as it is fraught with ambiguity and contradictory objectives. 'Appropriate measures will be developed […] to attract and retain international talent […] e.g. some foreign nationals graduating with PhDs at South African higher education institutions and graduates in scarce skills', but this 'must not be to the detriment of job opportunities for equally qualified and experienced South African citizens'. And then: 'Initiatives to attract and retain international talent from other African countries must be balanced against South Africa's obligation towards the development of the African continent'[15]. No indication is provided how these apparent contradictions will be resolved.

Given the critical role of the doctorate in the production and reproduction of knowledge, and of South Africa's established position on the continent as the prime destination for doctoral study, in this article we focus squarely on doctoral education and the prospects of South Africa as a PhD hub for the continent. We pose the following three questions: Is South Africa maintaining its position as a PhD hub for students from the rest of Africa? If so, what are the predominant factors influencing their selection of South Africa as a destination for doctoral study? And how do doctoral students view their prospects post-graduation?

In the sections below, the issues of reproduction and development, of mobility, and of diversity in relation to doctoral education are briefly explored in the context of Africa and South Africa. Reference is made to both the extant literature as well as to relevant policy. The purpose for doing so is to provide some context in which the main questions posed can be situated.

Literature review

Reproduction and development

The 'production' of doctorates is typically linked to its scholarly function of fulfilling the needs of the academic labour market.[16-18] Through doctoral education, future faculty are trained through a process of socialisation and research training. In Africa, and in other developing countries, the argument is made that there is an urgent need for doctoral graduates to take up academic posts to compensate for the relatively low numbers of academic staff with doctoral qualifications.[8] For example, only 43% of all permanent academics in South African universities had doctorate degrees in 2014[19], limiting the capacity of the system to supervise future doctoral candidates. Jørgensen[20] found that 33% of staff had a doctorate in 28 sampled institutions in Southern Africa in 2012, compared with 31% of staff in 29 institutions in Latin America and 49% of staff in 28 institutions in East Asia. The situation at several of Africa's leading research universities is no better.[21] A 2012 study in the UK showed that 58.4% of full-time staff had doctoral degrees.[22] More recent data for South Africa show that, by 2018, the proportion of permanent academic staff with doctoral degrees had increased to 47%[12] – well below the clearly ambitious 2030 target of 75% set by government[10,23].

In addition to securing the future of the academy, the doctorate is seen as providing high-level skills for the labour market.[16,14,24] The provision of skills links doctoral education to economic development by placing the emphasis on human capital. In South Africa, the initial thrust of policy was to 'address the local human capital requirements' because 'doctoral graduates [are] required to support a competitive knowledge-based economy'[25]. These and other policy statements suggest that national policymakers are concerned with addressing a human capital problem and view the doctorate as a mechanism for developing high-level skills. This should be read against emerging critiques of solutions premised on human capital such as arguments based on the idea that 'there's a strong ideological component behind the skills gap [argument]: it diverts attention (and policies) from the deep inequalities and market fundamentalism that created the unemployment crisis, and focuses on a fake skills gap that had nothing to do with the surge in joblessness'[26].

These two positions on the role of doctoral education suggest that the doctorate's knowledge production role is seen as the reproduction of the academy, while the skills produced contribute to a more productive economy. In this formulation, what is not evident is how the new knowledge produced by doctorates – during their training and in their future positions – makes an important contribution to development. As Backhouse[16] states in her assessment of the South African doctoral education: 'doctoral students are being trained to do research, rather than being engaged in doing research'. Based on a thematic analysis of 995 papers published between 1971 and 2012 on the topic of doctoral studies, Jones[24] identified six central themes covered by the corpus. None of the themes identified relates to the relationship between doctoral education as a knowledge endeavour and development. This may suggest a relatively recent recognition for doctorate's development value beyond the provision of high-level skills for the labour market.

South Africa's National Development Plan articulates the knowledge contribution of doctoral education to development. The first draft of the *National Development Plan: Vision 2030*[10] embraced the knowledge economy, declaring that 'knowledge production is the rationale of higher education'[10]. This is a radical departure from the traditional role of higher education in Africa.[27]

In a context of the increasing focus on the knowledge economy, the argument put forward is that if information and knowledge are the new electricity of the economy[4,28], then it is reasonable to assume that the university – as the main knowledge institution[29] – will become increasingly important, and that its apex training product, the PhD, will become more important and sought after.

The developmental contribution of doctorates places high value on attracting the best students to universities and to national university systems. The attraction of talent or 'talentism' is seen by some as superseding the importance of capital in modern economies.[30] The quest for talent finds expression in initiatives such as the UK's Research and Development (R&D) Roadmap which seeks to 'attract global talent, cut unnecessary bureaucracy and cement the UK as a world-leading science superpower'[31]. The government announced in June 2020, as part of the new Graduate Route, that international students who complete a PhD in 2021 can stay on in the UK for 3 years after study to live and work. Students who have successfully completed undergraduate and master's degrees will be able to stay on for 2 years after study. And a new Office for Talent is to be set up to make it easier for leading global scientists, researchers and innovators to live and work in the UK. Universities UK Chief Executive Alistair Jarvis said the announcement of extending the Graduate Route is a 'bold policy move which will increase the UK's competitive edge in the global competition for talented research students'[31].

Africa has been no exception in embracing the value of doctoral education, despite the fact that there is doubt as to whether any of the extractive-driven economies on the continent are knowledge-driven. This did not deter the African Union Commission Chair, Nkosazana Dlamini-Zuma, declaring that 'Africa must look at ways to train thousands more PhD students on the continent'[32]. Delegates from the African Research Universities Alliance concur that the role of African universities has

changed in the wake of the so-called Fourth Industrial Revolution, and that this requires universities to provide more doctoral training.[33]

Mobility

Knowledge migration, as the movement between universities as sites of knowledge, is inevitably a *relationship* between the students and scholars of different nation states. This is because the impetus for migrating is a factor of conditions in the country of origin, while the incentives put in place to encourage migration, postgraduate participation and the possibility of residency, are the prerogatives of the host nation. In other words, place matters because universities and the countries in which they are situated cannot be separated from the prevailing geo-politics and historically determined power relations.[34] Migration must therefore be understood in comparative terms.

This perspective supports the conceptualisation of contextually determined push and pull dynamics in determining whether students and scholars elect to migrate to countries other than their own for the purposes of education. The push–pull model suggests that students are 'pushed' from their home countries due to a lack of educational and other resources, and are 'pulled' to foreign countries to obtain better education and longer-term prospects.[35]

A 2014 survey of 1682 international students studying in South Africa found that pull factors among students from the rest of Africa (76% of the sample) include affordable tuition fees; proximity to home, affordable cost of living and social connections; the reputation and relative stability of higher education in South Africa; and the currency of qualifications obtained from South African universities in relation to employment prospects.[36] An increase in the number of academic staff from Africa may also play a role in attracting doctoral students from Africa.[8] A study on the retention, completion and progression rates of postgraduate students in South Africa[37] found that the three top reasons cited by non-South African doctoral students in choosing their doctoral degree qualification was the (1) academic reputation of the university, (2) their relationship with their (previous) academic supervisor, and (3) the availability of scholarships or bursaries. Kahn and Oghenetega[38] found that the high quality of academic programmes offered by South African universities, funding made available by South African universities and research funding agencies, and the availability of speciality programmes were the three main reasons for students from Africa selecting South Africa as a study destination.

At the same time, students from Africa face challenges that limit their mobility, such as the uneven application of study visa regulations.[39] Sporadic outbreaks of xenophobic attacks and campus instability following the rise of the Fallist student protests may also have tarnished the reputation of South Africa as an educational destination. Globally, the COVID-19 pandemic has unexpectedly halted the mobility of students. At the time of writing, movement remains barred to and from top doctoral destination countries such as the USA, Canada, the UK and Australia.

The brain drain perspective on migration and development takes mainly the perspective of the origin country into consideration. Migration is seen as detrimental to the development prospects of the country of origin. The brain circulation perspective[40] moves the discussion forward by suggesting that there are residual returns to the country of origin. From this perspective, the prospects of a better life post-graduation in another country may well influence the decisions of doctoral hopefuls and graduates when selecting a country of study and when choosing whether to remain in the country of their choice post-graduation. According to some, the economic and political factors that shape opportunities for graduates are often overlooked.[41,42]

Saxenian[40] describes 'new modern-day Argonauts' – technically skilled entrepreneurs who travel between Silicon Valley and their home countries as they seek to make their fortune. She shows how 'brain drain' has become 'brain circulation', and a driver for the development of formerly peripheral regions. The new Argonauts exploit their Silicon Valley networks and experience, and the ability to operate simultaneously across territories to identify market opportunities, locate foreign partners and manage cross-border business operations. Critically, Saxenian's new Argonauts offers a fresh perspective on how technology entrepreneurs build regional advantage to compete in global markets by arguing convincingly that the foreigner Argonauts have made America richer, not poorer. In the African context, there is anecdotal evidence that increasing numbers of doctoral graduates who pursued their studies abroad are returning home to Africa.[43] Kahn and Oghenetega[38] found that, based on a relatively small sample, 27.5% of 463 doctoral students from the rest of Africa remained in South Africa, 63.8% returned to their home countries on the continent, and 5.0% relocated to countries outside of Africa.

Despite an acknowledgement of the potential benefits of brain circulation, relatively little attention has been paid to the impact of knowledge migrants on the host nation, particularly in cases where regional hubs such as South Africa face their own post-colonial development challenges. The dilemma facing South Africa is, without the resources and system dynamics of Silicon Valley[14], to attract top doctoral students from the rest of Africa to contribute to the country's knowledge capacity, but not at the expense of developing local talent, thereby setting up a complex tension between underdevelopment and development[7,44]. Whether attempts to manage this tension result in brain gain, brain drain, brain circulation or some other form of brain economics, remains an open question.

Diversity

Diversity in higher education is both an established imperative and an ongoing concern as new issues emerge in changing contexts.[45] Research in the USA[46], Europe[47,48] and elsewhere continues to bring issues of diversity and the attendant complexities of its institutionalisation into the academic discourse. The argument for diversity is that the effectiveness, excellence and viability of universities depend on their diversity and, consequently, their ability to deliver on the promise of development, particularly in relation to the production of democratic societies.[45]

As student mobility globally and within Africa increases, so does the national diversity of students, including doctoral students, as they seek opportunity to further their studies at the limited number of research universities on the continent.[14,21] Diversity has been and remains a policy priority in South African higher education, including doctoral education, where it is subsumed under the dominant discourse of transformation.[14] The South African PhD Project launched by the National Research Foundation (in partnership with the then Department of Science and Technology) in 2007 seeks to intensify the strength and diversity of the higher education system by increasing the number and diversity of PhD graduates.[49] The most recent White Paper of 2013 sets out a vision of a transformed post-school system which it sees as integral to improving the economic, social and cultural life of South Africans: 'one that will be more equitable, much expanded and more diverse'[11]. While the White Paper does not provide clarity on what counts as diversity, it can be assumed that diversity includes cultural plurality, as well as a mix in the proportion of international students and staff at South African universities.

This policy position on diversity is one which is at loggerheads with other positions taken by government in relation to the regulation of the labour market and immigration.[8] The inconsistency at the level of policy – possibly intentional[39] – reflects perhaps the unresolved tension between knowledge-driven development that is global in its outlook and the persistent pressure to address local underdevelopment in South Africa. It is in this ambivalent context that the question of South Africa's position as a hub for doctoral study for the rest of Africa remains pertinent.

Methodology

Our approach is quantitative and relies on trend data to establish whether South Africa is maintaining its position as a doctoral hub for students from the rest of Africa. Data are critical because many unsubstantiated or partially unsubstantiated claims are made about the state of doctoral education in relation to the number and experiences of African doctoral students in the South African higher education system.[8,50] Our approach provides an update on previously published data on doctoral education in South Africa.[8,13,14,19,36]

We acknowledge Spencer-Oatey and Dauber's[51] argument about the limits of what they term 'compositional approaches' which seek to describe or measure internationalisation, that is, approaches that rely on 'parameters that focus exclusively on objectively countable characteristics'. To address this potential shortcoming, we conducted a survey of doctoral graduates to reveal some of the dynamics at play in the decisions taken by students in Africa who choose South Africa as a destination for their doctoral studies.

In this article, we draw on two sources of data. First, we rely on university performance data extracted from the South African Department of Higher Education and Training's Higher Education Management Information System (HEMIS) to identify trends with regard to African doctoral students in South Africa. We had access to the HEMIS microdata for students and staff from 2000 to 2019 which include demographic information such as nationality. The second source is qualitative data collected by SciSTIP (the DSI-NRF Centre of Excellence in Scientometrics and Science, Technology and Innovation Policy) from a large-scale survey targeted at doctoral students who completed their doctoral studies in South Africa between 2000 and 2019. The survey provided data for understanding post-graduation aspirations and mobility.

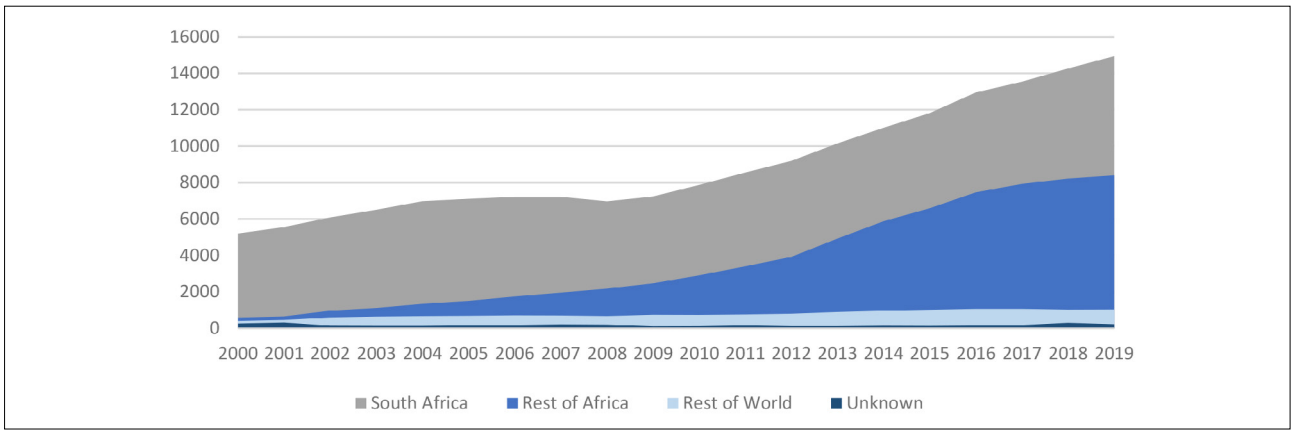
SciSTIP has developed a South African Doctoral Thesis Database which contains records of more than 95% of all doctoral graduates from South African higher education institutions between 2000 and 2019 for whom records exist on institutional repositories. The database includes details such as names, surnames, doctorate-granting institutions, years of degree completion and thesis titles of doctoral graduates. From the database, the research team identified email addresses from public domain sources for all entries who had graduated between 2000 and 2019. This selection produced email addresses for 18 578 graduates in the database. An electronic survey was successfully sent to 15 565 graduates, of which 6452 graduates responded (translating into a response rate of 41.4%). After cleaning the data, which included removing duplicates and excluding respondents who received the survey erroneously, a data set of 6438 observations was analysed using SPSS.

On the basis of four variables, namely gender, age at graduation, nationality and distribution of broad disciplinary fields, our sample of graduates was deemed representative of the population. We summarise the measures of representativeness in Table 1.

Of the 6438 respondents in our data set, 4024 (62.5%) indicated that they were South African citizens during their PhD studies, 1545 (24%) were from other African countries and 429 (6.7%) were citizens of countries outside of Africa.

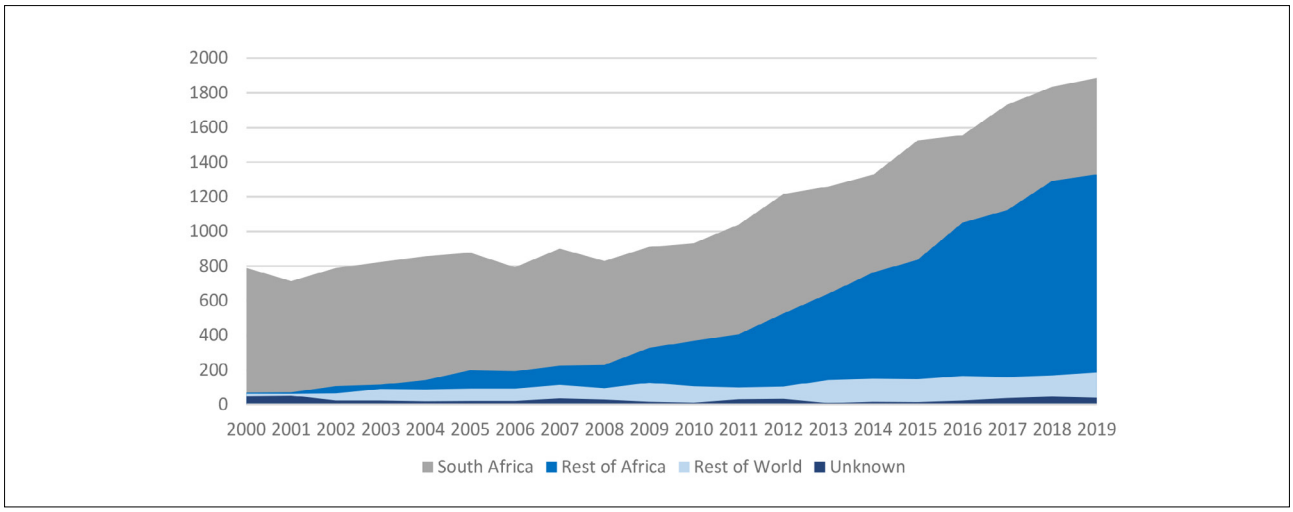
Table 1: Measures of representativeness of survey sample

Variable	Survey	HEMIS (2019)
Percentage female	45.3%	44.0%
Percentage South African	62.5%	54.8%
Percentage Science, Engineering and Technology fields	50.1%	49.9%
Age at graduation	39.8 years	40.5 years



Source: HEMIS microdata

Figure 1: PhD enrolments: Nationality of doctoral students enrolled in South African universities (*n*), 2000–2019.



Source: HEMIS microdata

Figure 2: PhD graduates: Nationality of doctoral graduates, 2000–2019.

Findings

Doctorates from Africa studying at South African universities

Figure 1 shows the steady growth in doctoral enrolments in South African universities, driven mainly by enrolments from South Africa and the rest of Africa. In terms of proportional representation, the data show that, in 2019, 4 out of every 10 doctoral students enrolled in South African universities were international students, with 1 out of 3 being from the rest of the African continent. Over the past 20 years, we have seen a dramatic increase in the number of doctoral enrolments from the rest of Africa with the percentage of African enrolments increasing from 9% in 2000 to 34% in 2019. However, over the past 4 years we see that the rapid growth in the proportion of African enrolments has slowed down and stabilised at around 34–35% of doctoral enrolments. Reasons for the plateauing of doctoral enrolments from the rest of Africa are unknown, and further study is needed to ascertain why this is the case. Possible hypotheses are unsettled university campuses in South Africa post-2015, changing funding opportunities, a tailing off in the number of master's graduates on the continent, and the attractiveness of new study destinations and platforms, to name a few.

When we look at the geographical breakdown of doctoral graduates as illustrated in Figure 2, we observe similar trends to that of enrolments, with significant shifts in the percentage of graduates from the African continent from 7% in 2000 to 39% in 2019, and South African graduates constituting 55% of graduates from South African public universities in 2019.

In Table 2, we show the changes in the percentage of doctoral enrolments from the top 16 feeder countries from Africa. We show the data only for 2012 and 2019 due to a change in coding classifications for nationality in HEMIS. With the exception of Zimbabwe, we see that the top four origin countries in 2019 are not part of SADC. Over the past 8 years, there has been little change in the origin countries of African doctoral enrolments, but with slight increases in the proportion of students from Zimbabwe, Nigeria and Ghana. We see slightly smaller contingents from Kenya, Tanzania, Uganda, Botswana and Malawi, but for the latter three countries the changes are negligible.

Table 2: PhD enrolments from the rest of Africa by country of origin (top 15 countries of origin, 2012 and 2019)[†]

Nationality	Percentage of total enrolments, 2012	Percentage of total enrolments, 2019	Proportional change
Zimbabwe[‡]	7.1%	10.2%	3.2%
Nigeria	4.4%	7.1%	2.7%
Ghana	0.7%	2.4%	1.7%
Ethiopia	2.0%	2.0%	0.0%
Kenya	2.3%	1.8%	**-0.6%**
Zambia[‡]	0.9%	1.1%	0.2%
Namibia[‡]	0.7%	1.0%	0.3%
Lesotho[‡]	0.9%	1.0%	0.1%
Uganda	1.2%	1.0%	**-0.2%**
Cameroon	0.8%	0.9%	0.1%
Eswatini[‡]	0.5%	0.8%	0.3%
Botswana[‡]	0.8%	0.7%	**-0.1%**
Malawi[‡]	0.8%	0.7%	**-0.1%**
DRC[‡]	0.6%	0.6%	0.0%
Tanzania[‡]	0.8%	0.4%	**-0.4%**

Source: HEMIS microdata

[†]Enrolments from South Africa: 65.5% in 2012; 60.8% in 2019

[‡]SADC countries

In Figure 3, we show the percentage of international students at South African institutions for 2005 and 2019 (we selected the first year after the mergers of South African universities). Unisa has consistently enrolled the largest proportion of international students, with more than half of their doctoral enrolments in 2019 constituted by non-South Africans. International students made up around 40% of doctoral enrolments at South Africa's bigger research universities (UKZN, UCT, UP and Wits).

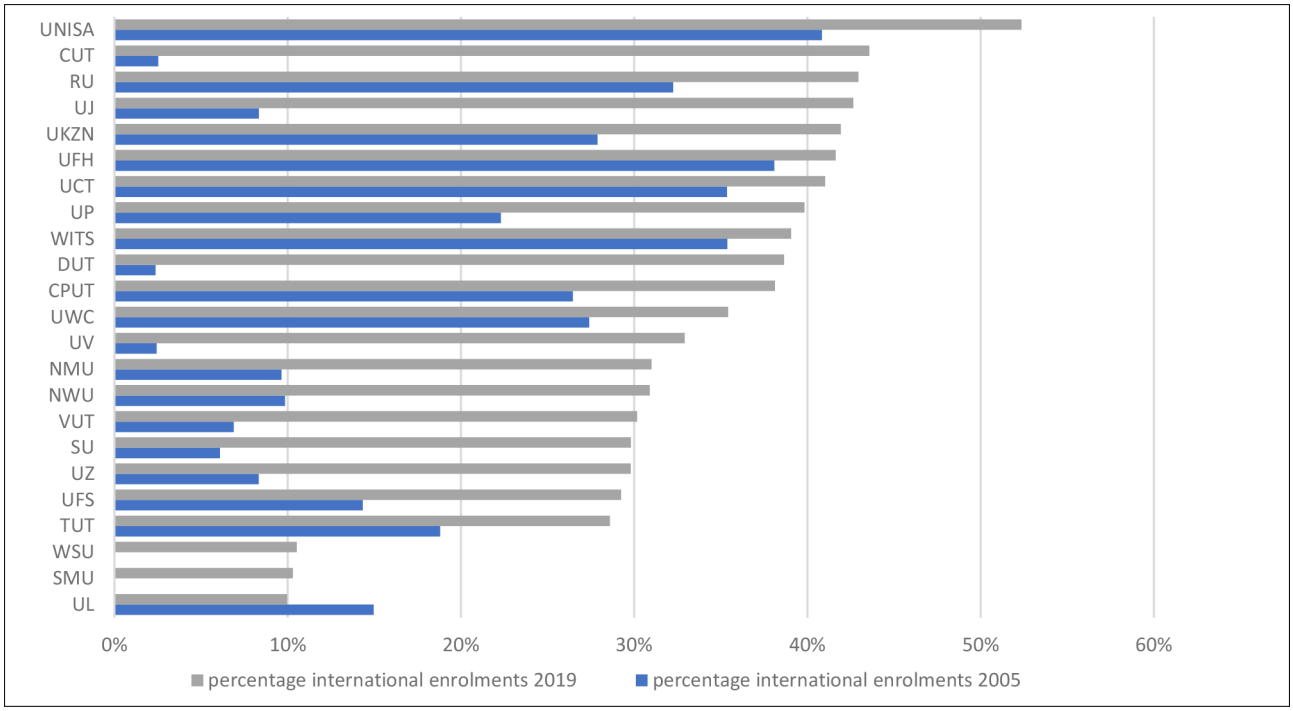


Figure 3: Percentage of international PhD enrolments by South African university in 2005 and 2019.

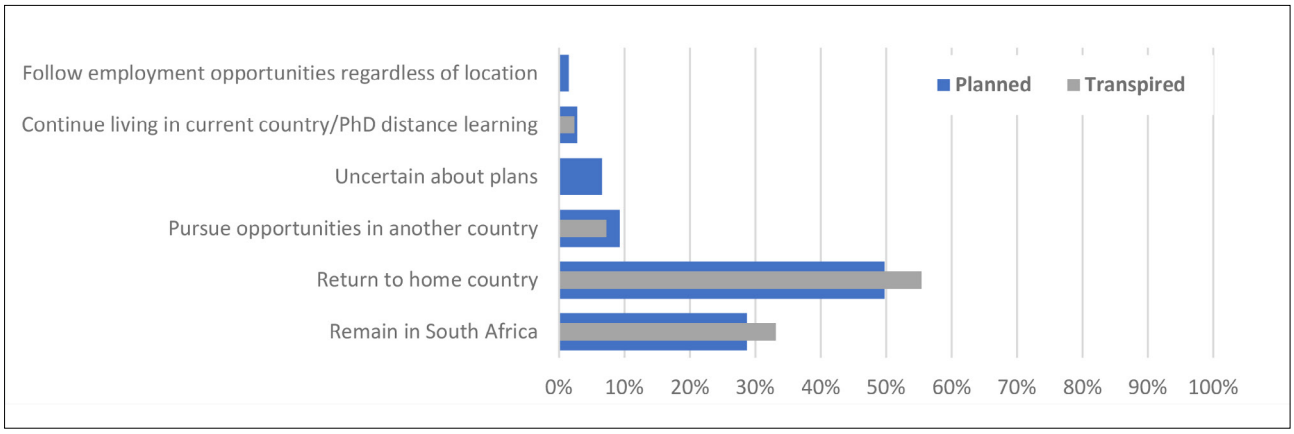


Figure 4: Mobility plans and outcomes of doctoral graduates from Africa.

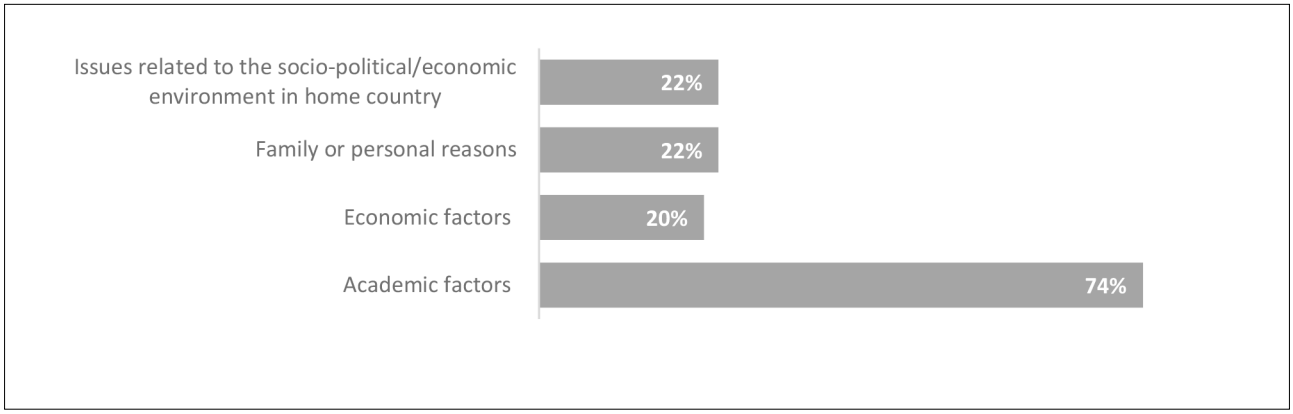


Figure 5: Reasons given by doctoral graduates from Africa for remaining in South Africa.

The prospects of doctoral graduates from Africa in South Africa

One of the aims of the doctoral graduate tracer study was to investigate the mobility patterns of doctoral graduates from South African universities. Survey respondents were asked to indicate what their plans were *during* their doctoral studies and to indicate what had transpired in the first year *after* graduation.

Figure 4 shows that it is evident that most doctoral graduates (50%) planned to return home after their doctoral studies. An additional 5% (55%) indicated that they had, in the end, returned to their home countries within the first year of graduating. A relatively high percentage of respondents also planned to remain in South Africa (29%), and it transpired that 33% of respondents ended up remaining in South Africa.

Of those respondents who remained in South Africa within the first year of completing their doctoral studies, the majority indicated that academic factors constituted the main reason for their decision (74%) (Figure 5). Economic and personal reasons were not cited as frequently as being the reason behind their decision to remain in South Africa.

Table 3 shows that when asked why respondents had left South Africa, family and personal reasons was the most common response (31%), although academic and economic reasons were also cited. Comparatively fewer respondents left South Africa due to issues related to visas (12%) or personal safety (8%). Few had commitments requiring their return to their home country (8%) or had no intention to stay in South Africa after graduating (3%).

Table 3: Reasons given by doctoral graduates from Africa for leaving South Africa after obtaining a doctorate

	n	%
Family or personal reasons	435	31%
Academic factors	265	19%
Economic factors	245	17%
Issues related to visa or residency in South Africa	170	12%
Issues related to personal safety (e.g. xenophobia)	116	8%
Employer/financial condition in home country	109	8%
Had no intention to stay	49	3%
Other reasons	21	1%

Source: HEMIS microdata

Discussion and conclusions

The description of South Africa as a doctoral hub on the continent, particularly for students from anglophone countries, remains an accurate one. Our findings confirm that, for the time being, South Africa remains an attractive destination for doctoral candidates from the rest of Africa.

When compared with the findings of Kahn and Oghenetega[38] that 64% of doctoral graduates returned home, our findings, based on a larger sample and one that includes doctoral graduates from all South African

universities, show that just over half (55%) of respondents returned home after graduating. Our findings show that 33% remained in South Africa compared to Kahn and Oghenetega's[38] finding that 28% remained in the country.

That most respondents indicated that they remained in South Africa for academic reasons squares with Kahn and Oghenetega's[38] finding that 17% of the 28% of doctoral graduates in their study remained in South Africa to pursue postdoctoral fellowships. It also supports the findings of previous studies that found the quality of South African university qualifications to be a strong motivator for selecting and remaining in South Africa.[36,37] This finding does, however, raise questions about whether doctoral graduates from Africa are using postdoctoral positions to mark time until other, longer-term opportunities to remain in South Africa arise. If this is the case, they are, to some extent, fulfilling a role as providers of skills for the academy, but they are less likely to make additional contributions to development if they are unable to secure a permanent university post, find employment in the labour market or engage in entrepreneurial activities. Moreover, the postdoc as temporary solution is under pressure, given indications that funding opportunities for postdocs from the rest of Africa are shrinking (see below).

In terms of reasons for leaving South Africa, respondents cited family reasons as the main reason for returning home. Issues related to personal safety (including xenophobia) and visa requirements played a relatively less important role in influencing respondents' decision to return home but nevertheless deserve attention as it could be argued that the fate of at least one in five doctoral students from the African continent may have been different under other circumstances – circumstances that are predominantly determined by the host government.

The percentage of doctoral graduates from Africa remaining in South Africa, and their reasons for doing so, should be interpreted in a context in which total doctoral enrolments at South African universities are growing, and in which the percentage of doctoral graduates from the rest of Africa in relation to those from South Africa is also increasing. While this aligns with the policy priority of a diverse student population and the retention of potentially valuable highly skilled labour to support the development of South Africa, an increase in doctoral graduates from the rest of Africa nevertheless has the propensity to be used by the South African government to support nationalist higher education and migration policies. The result is a thicket of policies that align selectively with the available data, and which leaves prospective doctoral candidates, doctoral graduates, university administrators and employees of doctoral graduates from African countries confused and frustrated. The consequence is brain circuity instead of brain circulation. In other words, policies related to the treatment of doctoral graduates from Africa, caught between different social, economic and political imperatives, result in contradictory and confusing signals, resulting in neither the circulation nor the draining or gaining of knowledge, but to a knowledge landscape characterised by indirection and unwanted intricacy.

For example, on the one hand, student enrolment and graduation data[52] show considerable progress in terms of equity. On the other hand, the focus on equity, or redress, results in the reservation of doctoral places and financial aid for black South Africans. The current policy of the NRF reserves 80% of funded places for black South Africans, and states that no foreigner may displace a South African citizen at a local educational institution.[53,54] Yet South Africa cannot reach the targets for doctorates set by government without the constant flow of candidates from the rest of Africa. An ARIMA (autoregressive integrated moving average) forecasting model of trends in doctoral outputs shows that the number of doctoral graduates from the rest of Africa is growing at a faster rate than that of South African graduates. If the trend continues, the primary driving factor in the production of doctoral graduates will be the influx of students from foreign countries. Based on the forecasting model for data between 2000 and 2017, there will be a point at which graduates from the rest of Africa surpass those from South Africa.[55]

Another confounding policy issue is that, in the national higher education database (HEMIS) managed by the Department of Higher Education and Training, South African racial classifications are assigned to foreigners. Such classification is both arbitrary and nonsensical. It also creates confusion when using the HEMIS data because non-South Africans are often included in reports disaggregated by race. This creates a misleading picture in terms of system-level transformation and diversity.

Since 2016, nationalism and xenophobia have become more prominent in the public discourse, leading to policies that favour South African black students. Confounding both positions is the policy of increased collaboration and preferential treatment for students from SADC countries when the data show that, in reality, most doctoral students from Africa do not come from the SADC region. Further analyses could include a focus on South African universities' internationalisation plans and agreements which undoubtedly influence the numbers of incoming students.

A positive development is that doctoral graduates from other African countries are more readily granted South African work permits if they meet the requirement of providing critical skills.[56] (These skills are all in technical and science areas; apparently there are no critical shortages in the arts and social sciences.) However, policies that rest on the principle of 'not displacing South Africans', while at the same time extracting scarce skills from other developing countries in Africa, attest to a lack of appreciation as to the scientific and developmental benefits of brain circulation. Such policies are also in contradiction with Africa-wide collaboration and the intentions of the recently signed free trade agreement.[57]

In 1996, the National Commission on Higher Education could not resolve the tension between equity (redress) and development, and this tension still underpins the contradictory policies of the South African government. Rather than supporting the circulation of knowledge for the benefit of all African countries, nationalist politics are stifling the contribution of South Africa's universities as key drivers of development across the continent. Brain circuity is a wasted opportunity.

Acknowledgements

We thank the Department of Higher Education and Training for making available the HEMIS data which were used for the analyses in this paper. This work is based on research supported by the National Research Foundation of South Africa (grant no. 91488). The survey of doctoral graduates forms part of a larger doctoral tracer study funded by the Department of Science and Innovation, with the support of the National Research Foundation, the Water Research Commission, the Department of Higher Education and Training, and Universities South Africa.

Competing interests

We have no competing interests to declare.

Authors' contributions

F.B.v.S.: Formulation or evolution of overarching research goals and aims; development and design of methodology; preparation and creation of the published work, specifically writing the initial draft and subsequent writing revisions; final approval of the version to be published. M.H.v.L.: Formulation or evolution of overarching research goals and aims; development or design of methodology: questionnaire construction, data cleaning and wrangling, survey and data analysis; writing of initial draft and subsequent revisions of manuscript; final approval of the version to be published. N.C. Formulation or evolution of overarching research goals and aims; writing of initial draft and revisions of manuscript; final approval of the version to be published.

References

1. Mouton J. Doctoral production in South Africa: Statistics, challenges and responses. Perspect Educ. 2011;29(3):13–29.
2. Nerad M. What we know about the dramatic Increase in PhD degrees and the reform of doctoral education worldwide: Implications for South Africa. Perspect Educ. 2011;29(3):1–12.
3. Cloete N, Mouton J, Sheppard C. Doctoral education in South Africa. Cape Town: African Minds; 2015. https://doi.org/10.47622/9781928331001

4. Castells M. Universities as dynamic systems of contradictory functions. In: Muller J, Cloete N, Van Schalkwyk F. Castells in Africa: Universities and development. Cape Town: African Minds; 2017. p. 35–55. https://doi.org/10.47622/9781920677923

5. Kehm BM, Teichler U. Research on internationalisation in higher education. J Stud Int Educ. 2007;11(3–4):260–273. https://doi.org/10.1177/1028315307303534

6. Knight J. Updating the definition of internationalization. IHE. 2003;33(2):2–3. https://doi.org/10.6017/ihe.2003.33.7391

7. Jansen J, McLelellan C, Greene R. South Africa. In: Teferra D, Knight J, editors. Higher education in Africa: The international dimension. Boston, MA: Center for International Higher Education & Accra: Association of African Universities; 2008. p. 387–420.

8. Cloete N, Sheppard C, Bailey T. South Africa as a PhD hub in Africa? In: Cloete N, Maassen P, Bailey T, editors. Knowledge production and contradictory functions in African higher education. Cape Town: African Minds; 2015. p. 75–108. https://doi.org/10.47622/978-1-920677-85-5

9. South African Ministry of Education. National Plan on Higher Education. Pretoria: Ministry of Education; 2001. https://www.dhet.gov.za/HED%20Policies/National%20Plan%20on%20Higher%20Education.pdf

10. South African National Planning Commission (NPC). National Development Plan 2030: Our future – make it work. Pretoria: SA Office of the Presidency, National Planning Commission; 2012. Available from: https://www.gov.za/sites/default/files/gcis_document/201409/ndp-2030-our-future-make-it-workr.pdf

11. South African Department of Higher Education and Training (DHET). White Paper for post-school education and training: Building an expanded, effective and integrated post-school system. Pretoria: DHET; 2013.

12. Bunting I, Badsha N, Van Schalkwyk F, Cloete N. Universities and the National Development Plan. Stellenbosch: CREST/SciSTIP; 2020.

13. Sehoole CT. Student mobility and doctoral education in South Africa. Perspect Educ. 2011;29(3):53–63. https://journals.ufs.ac.za/index.php/pie/article/view/1695

14. Cloete N, Mouton J, Sheppard C. Doctoral education in South Africa. Cape Town: African Minds; 2015. https://doi.org/10.47622/9781928331001

15. Levine MV. The skills gap and unemployment in Wisconsin: Separating fact from fiction. Center for Economic Development Publications. 2013;49, 18. Available from: https://dc.uwm.edu/ced_pubs/18

16. Backhouse J. Doctoral discourses in South Africa. Perspect Educ. 2011;29(3):30–39.

17. Moreau MP, Leathwood C. Graduates' employment and the discourse of employability: A critical analysis. J Educ Work. 2006;19(4):305–324. https://doi.org/10.1080/13639080600867083

18. Tomlinson M. Graduate employability: A review of conceptual and empirical themes. High Educ Policy. 2012;25:407–431. https://doi.org/10.1057/hep.2011.26

19. Breier M, Herman C. The PhD conundrum in South African academia. High Educ Q. 2017;17:352–368. https://doi.org/10.1111/hequ.12134

20. Jørgensen TE. CODOC: Cooperation on doctoral education between Africa, Asia, Latin America and Europe. Brussels: European University Association; 2012.

21. Cloete N, Bunting I, Van Schalkwyk F. The research university in Africa. Cape Town: African Minds; 2018. https://doi.org/10.47622/9781928331872

22. Grove J. Proportion of academics with PhDs. Times Higher Education. 2012 Nov 01. Available from: https://www.timeshighereducation.com/news/proportion-of-academics-with-phds-2012/421657.article

23. Bunting I, Badsha N, Van Schalkwyk F, Cloete N. Universities and the National Development Plan 2030: A report commissioned by Singizi Consulting for a review undertaken by the National Planning Commission. Unpublished report 2020.

24. Jones M. Issues in doctoral studies: Forty years of journal discussion: Where have we been and where are we going? Int J Doctoral Stud. 2013;8:83–104. https://doi.org/10.28945/1871

25. South African National Research Foundation (NRF). South African PhD project [brochure]. Pretoria: NRF; 2008.

26. Cloete N, Bailey T, Pillay P, Bunting I, Maassen P. Universities and economic development in Africa: Key findings. Cape Town: African Minds; 2011. http://www.africanminds.co.za/wp-content/uploads/2012/06/1340715376682271798.pdf

27. South African Department of Higher Education and Training. Policy framework for the internationalisation of higher education in South Africa. Pretoria: DHET; 2017.

28. Castells M. Communication power. Oxford: Blackwell; 2009.

29. Powell JJ, Baker DP, Fernandez F. The century of science: The global triumph of the research university. Bingley: Emerald; 2017. https://doi.org/10.1108/S1479-3679201733

30. Beach G. 'Talentism' is the new capitalism. The Wall Street Journal. 2014 Jul 17. Available from: https://blogs.wsj.com/cio/2014/07/17/talentism-is-the-new-capitalism/

31. O'Malley B. Foreign PhD graduates can stay on and work for 3 years. University World News. 2020 Jul 01. Available from: https://www.universityworldnews.com/post.php?story=2020070114003887

32. Namuddu K. Expanding and sustaining excellence in doctoral programmes in sub-Saharan Africa: What needs to be done? Report of a workshop convened by the National Research Foundation, South Africa, and Carnegie Corporation of New York; 2013 October 27–29.

33. Nakweya G. Call for Africa to generate its own knowledge and be free. University World News. 2019 Nov 28. Available from: https://www.universityworldnews.com/post.php?story=2019112608553030

34. Mamdani M. Decolonising universities. In: Halvorsen T, Orgeret KS, Kroyvel, R, editors. Sharing knowledge transforming societies. Cape Town: African Minds; 2019. p. 48–67. http://www.africanminds.co.za/wp-content/uploads/2019/10/Sharing_Knowledge_Transforming_Societies_web_ss.pdf

35. Altbach PG. Higher education crosses borders. Change. 2004;36(2):18–24. https://doi.org/10.1080/00091380409604964

36. Lee JJ, Sehoole C. Regional, continental, and global mobility to an emerging economy: the case of South Africa. High Educ. 2015;70:827–843. https://doi.org/10.1007/s10734-015-9869-7

37. Mouton J, Van Lill MH, Botha J, Boshoff N, Valentine A, Cloete N, et al. A study on the retention, completion and progression rates of South African postgraduate students. Stellenbosch: Centre for Research on Evaluation, Science and Technology; 2015.

38. Kahn M, Oghenetega J. Origins and destinations known: A tracer study of international African doctoral graduates from South Africa's universities. Ind High Educ. 2021;35(5):559–569. https://doi.org/10.1177/0950422221989645

39. Lee JJ, Paulidor K, Mpaga YA. Sliding doors: Strategic ambiguity in study visas to South Africa. Stud High Educ. 2018;43(11):1979–1992. https://doi.org/10.1080/03075079.2017.1296825

40. Saxenian AL. The new Argonauts: Regional advantage in a global economy. Cambridge, MA: Harvard University Press; 2007.

41. Bozalek V, Boughey C. (Mis)framing higher education in South Africa. Soc Policy Admin. 2012;46(6):688–703. https://doi.org/10.1111/J.1467-9515.2012.00863.X

42. Koen C. An analysis of research on graduate employment in South Africa. In: Koen C, editor. Higher education and work: setting a new research agenda. Pretoria: Human Sciences Research Council; 2006. p. 1–30.

43. Alemdjrodo R, Omanga D. Lessons for universities as diaspora scholars return home. The Standard. 2019 Jan 12. Available from: https://www.standardmedia.co.ke/article/2001309199/lessons-for-universities-as-diaspora-scholars-return-home.

44. Rodney W. How Europe underdeveloped Africa. Dar es Salaam: Tanzania Publishing House; 1970.

45. Smith DG. Diversity's promise for higher education: making it work. Baltimore, MD: Johns Hopkins University Press; 2020.

46. Kwak N, Gabriela Gavrila S, Ramirez FO. Enacting diversity in American higher education. In: Christensen T, Gornitzka Å, Ramirez F, editors. Universities as agencies: Reputations and professionalization. Cham: Palgrave Macmillan; 2019. https://link.springer.com/chapter/10.1007/978-3-319-92713-8_8

47. Brooks R. Diversity and the European higher education student: Policy influencers' narratives of difference. Stud High Educ. 2020;45(7):1507–1518. https://doi.org/10.1080/03075079.2018.1564263

48. Claeys-Kulik A-L, Jørgensen TE, Stöber H. Diversity, equity and inclusion in European higher education institutions: Results from the INVITED project. Brussels: European Universities Association; 2019. Available from: https://www.eua.eu/resources/publications/890:diversity,-equity-and-inclusion-in-european-higher-education-institutions-results-from-the-invited-project.html

49. South African National Research Foundation (NRF). SA PhD project [webpage on the Internet]. No date [cited 2021 May 28]. Available from: https://www.nrf.ac.za/saphd

50. Teferra D. Manufacturing – and exporting – excellence and 'mediocrity': Doctoral education in South Africa. S Afr J High Educ. 2015;29(5):8–19.

51. Spencer-Oatey H, Dauber D. Internationalisation and student diversity: How far are the opportunity benefits being perceived and exploited? High Educ. 2019;78:1035–1058. https://doi.org/10.1007/s10734-019-00386-4

52. Van Schalkwyk FB, Van Lill MH, Cloete N, Bailey T. Transformation impossible: Policy, evidence and change in South African higher education. High Educ. Forthcoming 2021. https://doi.org/10.1007/s10734-021-00687-7

53. South African National Research Foundation (NRF). Black Academics Advancement Programme (BAAP) framework. Pretoria: NRF; 2020. Available from: https://www.nrf.ac.za/sites/default/files/Black%20Academics%20Advancement%20Programme%20%28BAAP%29%20Framework%202021%20Call%20Final.pdf

54. South African National Research Foundation (NRF). Freestanding, innovation and scarce skills postdoctoral fellowships framework. Pretoria: NRF; 2021.

55. Mouton J, Basson I, Blanckenberg B, Boshoff B, Prozesky H, Redelinghuys R. The state of the South African research enterprise. Stellenbosch: DSI-NRF Centre of Excellence in Scientometrics and Science, Technology and Innovation Policy; 2019.

56. Staff writer. Here is a list of critical skills needed in South Africa. Business Tech. 2021 Feb 18. Available from: https://businesstech.co.za/news/government/469194/here-is-the-list-of-critical-skills-needed-in-south-africa/

57. Kende-Robb C. 6 reasons why Africa's new free trade area is global game changer. World Economic Forum. 2021 Feb 09. Available from: https://www.weforum.org/agenda/2021/02/afcfta-africa-free-trade-global-game-changer/

AUTHOR:
Johann Louw[1]

AFFILIATION:
[1]Department of Psychology, University of Cape Town, Cape Town, South Africa

CORRESPONDENCE TO:
Johann Louw

EMAIL:
johann.louw@uct.ac.za

DATES:
Received: 07 Feb. 2021
Revised: 19 May 2021
Accepted: 23 May 2021
Published: 29 Sep. 2021

HOW TO CITE:
Louw J. A thesis embargoed: Personnel research and ideology in South Africa after World War II. S Afr J Sci. 2021;117(9/10), Art. #9512. https://doi.org/10.17159/sajs.2021/9512

ARTICLE INCLUDES:
☒ Peer review
☐ Supplementary material

DATA AVAILABILITY:
☐ Open data set
☐ All data included
☐ On request from author(s)
☐ Not available
☒ Not applicable

EDITORS:
Chrissie Boughey
Jenni Case

KEYWORDS:
history, personnel research, African civil servants, apartheid, NIPR

FUNDING:
National Research Foundation of South Africa (grant no. 119117)

A thesis embargoed: Personnel research and ideology in South Africa after World War II

Ten years after the conclusion of World War II, the Department of Native Affairs of the National Party government of South Africa sponsored research into the selection of African civil servants. The study was conducted by Rae Sherwood, under the auspices of the National Social Research Council, and the National Institute for Personnel Research. In 1960, Sherwood submitted the work to the University of the Witwatersrand to obtain a PhD degree. Two government departments objected to the award of the degree. In this paper, I recount the history of the research, explaining that the acceleration of the apartheid project between 1948 and 1961 played a significant role in the controversy that developed. The paper furthermore illustrates the difficulties faced by social scientific research under repressive political conditions, and the need for a more nuanced view of the psychological research of the National Institute for Personnel Research in South Africa at the time.

Significance:

- The history of South Africa's research organisations has been of interest for a long time. This study recalls the history of an unknown chapter in the history of the National Institute for Personnel Research, based on a PhD submission kept under embargo in the archives of the University of the Witwatersrand. The study was methodologically sophisticated, rich in data, but controversial in its findings, at least as it reflected on the policies pursued by the apartheid government after World War II. It adds another contextual element to the type of work conducted by the Institute.

Introduction

In July 1960, Rae Sherwood submitted a PhD thesis, 'The African Civil Servant – A Socio-Psychological Study', to the University of the Witwatersrand.[1] In an interview I conducted on 23 August 1983 with Dr Sherwood in London, she mentioned that initially the National Party government refused permission for the research to be published as a PhD and that she had to reapply in the 1970s to obtain permission for the work to be considered for the degree. The degree was awarded in 1973, but she reported that even then the government placed it under an embargo. In 1995, I asked the University library whether the thesis was available for perusal, only to be informed that the thesis was not for consultation.[2] In 2008, when I visited the University archives, a staff member showed me the thesis, sealed in a large brown envelope. In 2018, I formally asked the University to lift the embargo, as it was an absurdity to have an apartheid-era thesis treated like this. After consultation with and approval from the executors of Dr Sherwood's estate in 2018 and 2019, the University formally made the thesis available.

Before embarking on the PhD, Sherwood had completed her first degree at the University of the Witwatersrand in 1940 and a master's degree in human development at the University of Chicago in 1948. She also received training at the Tavistock Institute in London in 1946. The family left South Africa in 1959 for the USA and settled in the United Kingdom in 1961. After the Sharpeville shootings of 1960, a large number of professionals and semi-professionals left South Africa – an exodus often characterised as South Africa's 'brain drain'. The fate of her brother, Rusty Bernstein, demonstrated vividly the risks for those who opposed the government at this time: between 1956 and 1958 he appeared in the famous Treason Trial, and in 1960 he was detained without charge for almost five months during the post-Sharpeville state of emergency. In the UK, Sherwood held positions at the London School of Economics, Brunel University, the Tavistock Institute, and the University of Sussex. While at Sussex, under the auspices of the Columbus Centre, she conducted research, based on psychoanalytic interviews, on personality and racial attitudes of three British families. The work culminated in her major publication, *The Psychodynamics of Race*.[3]

Although the research study on African civil servants was only one among hundreds of studies conducted under the auspices of the South African National Institute for Personnel Research (NIPR) in the 1950s and 1960s, its history merits uncovering. For a start, a work that has been buried in the archives for so many years upon the instructions of the apartheid government justifies a closer look, for this reason alone. In addition, it may provide interesting information about personnel research conducted just after World War II.

Luruli and Mouton[4] gave a useful overview of the early history of research institutes and funding in South Africa. The NIPR (initially called the National Bureau for Personnel Research) was established in April 1946 within the Council for Scientific and Industrial Research (CSIR), with Simon Biesheuvel as director. Its goal was to produce scientific research in personnel selection and management in a South African economy that was developing rapidly after World War II. The Institute's contract work included research for the mining industry, the Defence Force, and public service departments like the Post Office and the SA Railways and Harbours.

The present paper addresses two questions: What were the events that constituted the controversy surrounding the thesis? and What in the thesis elicited such a strong reaction from the government of the time?

Trouble with the thesis

Sherwood recalled that the National Social Research Council (NSRC) made a grant available to study African civil servants, with an eye on selecting 'men who would be able civil servants, hardworking and honest and honourable' (Sherwood R 1983, personal communication, 23 August). The records of the NSRC in the National Archives show that, in May 1954, it approved a request from the Secretary of the Department of Native Affairs (DNA), Dr W.W.M. Eiselen, to conduct research on the selection and efficiency of the 'Bantu civil servant'.[5]

A Supervising Committee was approved by the Minister of Education, Arts and Science, which included the Director of the Bureau of Educational and Social Research as chair (S. Pauw, later replaced by P.M. Robbertse), P.A.W. Cook and A.C. Myburgh, both from the DNA, and S. Biesheuvel of the NIPR. The records of the Supervising Committee[5] intimated that they were looking for an experienced researcher. Sherwood's master's degree from the University of Chicago, a good research record, and the fact that she represented the Union government at a UNESCO conference in Mexico, made her a good candidate to conduct the research in the eyes of the Committee. She was appointed as a part-time Senior Research Officer at the NIPR in July 1954, with the responsibility to lead the research under the direction of this Committee.

It is difficult to establish from the existing documentation why the DNA sought to improve the selection and training of African civil servants. The only direct evidence for the DNA's motivation is contained in an untitled document from April 1954.[6] It stated (translated from Afrikaans) that the problem in South Africa and indeed in Africa was that 'Native personnel' were often appointed with no consideration of their intellect or aptitude for the tasks they had to fulfil. This had a negative impact on their work efficiency and prevented them from making a useful contribution to their developing societies. It therefore would be worthwhile if it were possible to identify 'gifted Natives' who could accept responsibility and who had an aptitude for their jobs, argued the unidentified author (compare Sherwood's statement above about able, hardworking, and honourable civil servants). The Department therefore proposed the development of psychological tests, according to this document, which were designed to determine what aptitudes were required for different jobs in the civil service. The DNA clearly was looking for a practically useful outcome: the delivery of psychological tests for the selection of African civil servants. The NIPR, with its extensive experience of developing such tests for the classification of African workers in the gold mining industry, was a logical choice to conduct the research.

The broader political context may provide further pointers as to the DNA's motivation for this research. The National Party in the early 1950s started to address the challenge posed by urbanisation and administration of black South Africans to its apartheid vision. The DNA, with a wide range of responsibilities covering African labour bureaux, housing, immigration, and tax collection, played an essential role in controlling this challenge. During the 1950s, the DNA grew from a relatively small department to 'a largely autonomous "state within a state"'[7(p.30)], with a staffing complement of more than 3000. Hendrik Verwoerd, who later would become Prime Minister, was appointed as the Minister of Native Affairs in 1950, to reorganise the Department into a much more efficient unit. Posel[8] has argued that as a social scientist himself, Verwoerd was one of the most assertive champions of the powers of science as the basis for informed planning, and the need for bureaucrats to administer and control the urban African population.

At the start of the study the Supervising Committee interpreted its mandate quite broadly:

> *every factor which might be supposed to have a bearing on the efficiency of the Bantu civil servant … would be investigated … Before practical recommendations involving the selection, training and management of these workers could be made, it was essential that a very broad foundation of understanding of the social, cultural and psychological factors involved in work attitudes and job performance be established.*[9(p.1)]

The motivation for framing the research in this way, broader than the DNA requested, cannot be ascertained from the documentation, but it clearly created an opportunity to extend the project beyond its narrow focus on selection and testing. The Supervising Committee, in broadening the scope of the study in this way, created essentially two studies: one aimed at developing tests for selecting African civil servants, requested by the DNA; and a more politically sensitive one in which the overall functioning of African civil servants in the DNA would be explored. It is not clear who convinced the Supervising Committee to re-direct the study along these lines, but, in her interview, Sherwood suggested that she had argued for such a broader study, and that the Committee had agreed to it. As it turned out, the development of psychological tests received much less attention than the social, cultural, and psychological factors mentioned above. For example, the NIPR published only an interim report on the 'selection and work efficiency of Bantu public servants' as late as December 1961.[5] The DNA was less than impressed with this situation, as described below.

In the progress report for the period October 1956 to October 1957[10], two project reports were included: a 43-page report on 'The Bantu Clerk: A study of role expectations', and 'Motivational analysis: A comparison of job attitudes among African and American professional and clerical workers'. The Committee recommended that the former be published in the *Journal of Social Psychology*[11], while the second one was presented at the 1957 conference of the South African Psychological Association[12]. The report noted that approval to publish was obtained from Eiselen as Secretary of the DNA, indicating that the Department was at least informed about publications from the study.

Sherwood's intention to use the material to write a thesis was never a secret. In September 1955, the Supervising Committee noted that she requested to use material not covered by the report to the DNA for a doctoral thesis, but preferred that 'she should leave the matter over until she had completed the final report, and that a request containing details of the aspects involved would receive sympathetic consideration and would be submitted with the Committee's recommendations to the Department of Native Affairs'.[13(p.2)]

The project drew to an end by the early 1960s. Sherwood handed in the thesis in July 1960, without informing the Supervising Committee. The University records show that it was 'based on a research project carried out in the National Institute for Personnel Research under the auspices of the National Council for Social Research'.[14(p.1)] In October 1960, the Secretary of the Department of Education, Arts and Science (DEAS) informed the Registrar of the University that it had come to their notice that Ms Sherwood had submitted a thesis for a PhD.[15] It is not clear how it came to their notice; the Secretary wrote only that it was 'by accident'. The Registrar was asked whether it concerned the selection and efficiency of Bantu civil servants. The letter also pointed out that Sherwood was employed by an ad hoc committee (the Supervising Committee) of the NSRC, who had not given permission to use the results for her own purposes. Not only that, but the report on the research had not been completed and submitted.

Sherwood replied that it was her understanding from the beginning that she would be permitted to submit part of the work for a thesis, and that her request was supported by the Committee.[16] She had regularly discussed it with Biesheuvel as her supervisor and had applied to the University for admission to the PhD with his approval. She remarked that the thesis she submitted was very different from the report to be submitted to the NSRC, in that the report was oriented toward the practical problems of personnel selection, while the thesis was concerned with exploring a number of theoretical propositions as 'a contribution to science'. Furthermore, the interim report was completed in May 1960 while the thesis was submitted in July 1960.

Biesheuvel also wrote to Robbertse[17] as Chair of the NSRC and indicated that he too believed that Ms Sherwood had received permission to proceed with the thesis, and that she had always acted with his knowledge and approval. Biesheuvel reiterated that Sherwood's thesis was not about the selection of clerks, but on the development of personality. A month later, however, after scrutinising the documentation, Biesheuvel acknowledged to Robbertse that he might have been under

a wrong impression and that Sherwood needed prior permission from Council to submit the thesis.[18]

The summary of the history of the project showed that Sherwood signed her contract on 5 October 1954. The contract stipulated, inter alia, that:

> *all information collected by the Grantee during the course of the investigation shall be considered confidential and shall not be divulged or in any way made public without the written consent of the Supervising Committee, provided that if the Grantee wishes to use data collected by her during the investigation for the purpose of presenting a thesis to a university she may apply to the Supervising Committee for such permission.*[5(p.2)]

When the DEAS received the thesis from the University, they sent it to Eiselen, asking whether the thesis would embarrass the DNA (by that time Eiselen was no longer Secretary of the DNA). Eiselen's report[19] gave the thesis serious consideration, providing an accurate and fair summary. He regarded much of the thesis as quite technical, with '*duistere vakterminologie*' (dense jargon), and therefore unlikely to cause much harm. However, the main findings and conclusions were stated in very clear language (see the discussion on the results of the thesis below), which could be grasped quite easily in reviews and in the press. It may thus be, he concluded, that the Department had commissioned research to find methods to improve the efficiency with which the African clerks delivered services to their communities, but instead would be castigated for its role in administering 'unjust' government policies of separate development.

The Secretary of DEAS subsequently informed the University that permission could not be granted.[20] Apparently, the Department of Bantu Administration and Development (DBAD), as the DNA was called by then, at this stage played no role in the decision. The Secretary for the DBAD wrote to the DEAS only on 7 September 1961, noting that it had come to their attention that Sherwood had submitted a thesis, and asking for a copy.[21] On 15 June 1962, the DBAD wrote that it could not give permission for the work to be submitted – nine months after the DEAS had reached the same decision.[22]

A few months later, the Secretary for the DBAD gave reasons to the DEAS for the recommendation.[23] It claimed that the data on which Sherwood reached her conclusions were incomplete and did not represent the views of the ordinary 'Bantu civil servant'. It was furthermore claimed that the findings were already dated, as many senior positions had been created in the civil service for 'Bantus' since the thesis had been completed (the letter cited no evidence to support this claim). Thus, stated the DBAD, releasing the thesis would not present an accurate picture of the situation and may in fact interfere in the Department's efforts to improve the situation. Briefly put, releasing the thesis would not have had any positive outcomes.

By September 1962, therefore, the fate of the thesis was settled: both government departments involved in the study refused permission for the thesis to be considered for a degree. In February 1962, the Secretary of the DEAS warned Sherwood in no uncertain terms, that 'Should you publish the material or make use of it in any way without the approval of the National Council for Social Research, the matter will be referred to the Attorney General…'[24]. The University of the Witwatersrand informed Sherwood that they could not accept the thesis without this approval.

As far as the original request from the DNA was concerned, in July 1961, the Secretary of the DEAS asked the President of the CSIR for copies of the complete set of psychological tests on the selection of African civil servants, as well as for a manual with information about norms, reliability, and validity. The letter stated further that the tests that had been sent to them (the NSRC files in the Archives contain nothing on this) were practically 'useless' and warned that no further funding would be forthcoming.[25]

In December 1961, the Supervising Committee submitted an interim report on the development of these tests.[26] It provided a summary of the research conducted, but with no mention of the developments around the thesis. It admitted that the final practical outcomes were disappointing, to say the least. According to the report, it was not yet possible to devise a set of personality tests that could be used for selection of African public servants, while the results of the tests of ability that were tried were not promising. The test results showed weak relationships with efficiency criteria such as productivity and job satisfaction considered in the study. Nevertheless, a preliminary battery of tests was finally constructed, named the 'Normal Battery'. It contained an omnibus set of tests: of intelligence, reading comprehension, vocabulary, spelling, and computation. Although not fully standardised, both Afrikaans and English language versions were available.

The response of the DBAD to these interim results was no less severe than the views of the DEAS a year before. The Secretary of the DBAD informed the Secretary of the DEAS that it was impossible to make use of the tests developed by the NIPR. The two versions of the Normal Battery, as well as the Interim Report, were thereby returned, and as far as the archival records go, this was the end of that part of the project.[27]

The thesis, and the award of a doctorate, not surprisingly, were very important to Sherwood, and she made two further attempts to get the government departments to overturn their refusals. In October 1965, she again requested the University to approach the DEAS. The Registrar wrote to the Secretary of the DEAS, reiterating the University's preparedness to regard the contents of the thesis as secret, to be released only to the examiners, and not to be placed on the library shelves.[28] Robbertse approached Eiselen again, this time to ask whether in his opinion the thesis should go forward for examination.[29] Eiselen replied within a week, stating that the thesis may very well be unacceptable to the DBAD, as it provided a twisted presentation of government policy.[30] Although he did not want to comment on the academic merit of the thesis, he nevertheless remarked that the work did not show the necessary objectivity required from a serious academic researcher. This observation notwithstanding, he expected that the thesis most likely would be accepted as adequate for a doctoral degree, and if the University could keep it secret, it could be submitted.

Eiselen's prediction about the DBAD turned out to be accurate, as it informed the Secretary of the DEAS four months later that it could not approve submission.[31] The DEAS waited until March 1968 to inform the University that the thesis still could not be accepted for submission towards a doctoral degree.[32]

Sherwood, however, was not ready to give up. Five years later she approached the University again, whereupon the Faculty of Arts appointed an ad hoc committee to consider the request.[14] (As the records of the National Archives are silent about these events, I relied upon the University archives for a description of what transpired in 1973). The minutes of this committee's meeting show that when the thesis was submitted in 1960, three examiners were appointed before the process was halted: Professors Simon Biesheuvel and I.D. MacCrone of the University of the Witwatersrand, and Kurt Danziger of the University of Cape Town. The reports were generally very favourable, although some revisions were required. Then, on '14th September 1961, the Secretary for Education, Arts and Science wrote to confirm that permission would not be granted to the University to "consider" the thesis submitted by Mrs Sherwood for the Ph.D. degree'[14(p.2)]. The minutes show that only when Sherwood and the University re-submitted their representation to the DEAS in 1973, was permission given, on 20 September 1973, that the work could be considered for a doctoral degree. The conditions, however, were that the work should be treated in the strictest of confidence, and was 'not to be reproduced or placed in the University library where general readers could have access to it'[14(p.3)]. These were exactly the conditions the University was prepared to accept eight years previously to enable the thesis to proceed to examination.

At its meeting on 1 October 1973, the ad hoc committee considered the original examiners' reports and subsequent revisions of the thesis. The head of the Department of Psychology, Prof. Jack Mann, was requested to consider the thesis, the examiners' reports, and the candidate's revisions, to see whether Sherwood adequately addressed the concerns

of the examiners. Upon receiving Mann's favourable report[33], the committee decided, given the exceptional circumstances surrounding the thesis, that the degree be awarded, but under the conditions stipulated.

In conclusion, it is fair to say that the original purpose of the study, to develop psychological instruments that could be used in selecting African clerks for the DNA, came to an embarrassing end. The Interim Report referred to above arrived six years after the study was launched and was brief and cursory. Apart from the two tests, with no data on norms and no manual, there was not much that the Department could use, as the authors of the Report in fact acknowledged themselves. By contrast, Sherwood's thesis consisted of 221 single-spaced pages, in which she presented a huge amount of qualitative and quantitative data, together with carefully developed arguments. It is difficult to avoid the conclusion that the thesis part of the project superseded the less ambitious aims that the DNA originally had in mind. As indicated earlier, the choice of the Supervising Committee to include social, cultural and psychological factors involved in work attitudes and job performance in the research set the research project on this path.

The second part of the paper explores the thesis itself, to find out why it was so troublesome and why the question arose whether its contents would be embarrassing to the DNA.

Results and conclusions of the thesis

To establish 'the very broad foundation' that the Supervising Committee envisaged, Sherwood drew a large sample (N=392) of African clerks, teachers, social workers, and nurses together to study middle-class social attitudes and values. The group of African clerks comprised all employees of the DNA in Johannesburg (57); a random sample of African clerks working for the DNA in Pretoria (26); all African clerks employed by the Johannesburg City Council (102); and an informal sample of similar employees of private firms in Johannesburg (20). Chapters 3, 4 and 5 of the thesis describe the results of this part of the study.

The 57 African men who were employed as clerks in the Johannesburg office of the DNA provided the data for the rest of the thesis (Chapters 7 to 10). The chapters cover themes such as personality, work ideology and motivation, role expectations, and work setting. 'Work setting' included supervision, and the study therefore also included 53 white employees of the DNA who supervised these 57 African clerks. The inclusion of white supervisors could not have been accidental. The NIPR, and Biesheuvel in particular, was alert to the inherent conflict between African workers and their white supervisors, originally arising from the NIPR's work within the gold mining industry.[34]

The term 'role conflict' perhaps best summarises the overall conclusion of the study, where conflict occurred between the roles demanded from African clerks by the official policies of the South African government, white supervisors within the Department, and the African public that the clerks served. Government and the supervisors placed 'a minimum emphasis upon African ability, self-determination, the exercise of initiative, and the full expression and development of African potential'[1(p.209)]. The DNA instead expected the clerks to be compliant, dependent, and ingratiating to authority. The white supervisors reproduced this conflict, as they gave high ratings to those clerks who were passive and dependent, irrespective of how efficient and productive they were in the organisation. As a result, the men had many complaints and sources of dissatisfaction in their work situation. A conclusion that 'The research findings indicate that the civil servant clerks complain more often of grievances in connection with apartheid and the colour bar than they do of any other frustrations affecting them in their work situation'[1(p.210)] must have been especially unpalatable to the DNA.

The role conflict furthermore extended beyond the boundaries of the organisation, to their relations with the African public, Sherwood argued. Clerks in the DNA dealt almost exclusively with other Africans, in activities such as assisting Africans with completing the DNA's forms, receiving payments, registering official statistics, taking statements on various disputes, and acting as interpreters during interviews or in court. Because they had to enact hated government policies against the interests of the African public, they became the focus of the negative feelings and rejections of fellow Africans. Indeed, Sherwood pointed out that the data reflected the social situation in 1956, in which Africans regarded this Department with great hostility, and African clerks who worked for the Department were regarded as 'traitors' to the people, which caused them, at the personality level, acute conflict, and strain.

The wider implication was that working for the Department would be unattractive to certain types of personality, and people with these personality types would therefore not select it as a place of work. At the same time, even if people who did not exhibit these ideal personality types did apply, they were not selected or retained. Sherwood predicted that the conflict that these clerks experienced would only worsen because the political climate was changing for the worse in South Africa. In summary,

> *The evidence from this analysis is clear: the present role structured for the African civil servants both limits and impedes the efficiency of these employees, and hence of the organisation itself, and creates frustration, conflict, and constriction in the individuals who occupy these roles.*[1(p.212)]

This rich data-gathering strategy demonstrated the importance of taking social, cultural and psychological factors into account, as it showed:

> *that work behaviour can be understood only with reference to the wider social and institutional complex of society. African work problems and their analysis have inevitably involved the study of African and white roles in the society as a whole and the pattern of their inter-relations. The organisation within which men work and function is not a closed social system, unaffected by the wider society, but is, in microcosm, a study of that society itself'.*[1(p.220-221)]

In the final paragraph, Sherwood explicitly drew out the implications of her study, not just for the African civil servants, but for black people in South Africa in general. This must have been the kind of statement Eiselen had in mind when he considered parts of the thesis as twisted presentations of government policy:

> *Where a society such as South Africa relegates subordinate status to a group of its citizens, and where discriminatory legislation affects their political, economic, and social lives, this study has shown that such subordinate groups cannot use their full potentialities constructively or creatively, and that their desire to contribute in full measure to the economic processes of that society will be undermined and weakened. To structure a role for a group of people so that their subordination is more important than their efficiency or their feelings of job satisfaction is to sacrifice not only their productive labour but also to undermine their morale and to build up hostilities within the group so that work behaviour comes to be an arena for the working through of their frustrations and resentments.*[1(p.221)]

The paragraph quoted at some length above is as eloquent an argument against apartheid, based on social science data and theory, that one could find in 1960s South Africa. Even the use of the word 'citizens' was problematic, as black people belonged in their 'homelands' and not 'South Africa'. It also provides an explanation for why the results of the study were so troublesome, especially if seen against the background of a rapidly deteriorating relationship between the NIPR and the National Party government in the 1950s.

The NIPR in the 1950s was an anomaly among semi-government institutions, according to Coupe[34]. It increasingly ran into difficulties with the National Party government as it did not fall in line with apartheid ideology. For example, one NIPR staff member, Yette Glass, was refused

a visa to present at an inter-African labour conference, because of her refusal to adopt the term 'Bantu' in her work.[34(p.61)] Sherwood herself avoided the term 'Bantu' in the thesis, except in the questionnaire for white supervisors. She explained in a footnote that 'The word Bantu was used here in order not to antagonise the white subjects concerned' (underlining in original).[1(Appendix B2)] Biesheuvel[35] recalled that he too was barred from attending at least two conferences on African labour and reminisced about the 'harassment' of the NIPR, which led to contracts from government virtually ceasing. He resigned from the NIPR in 1962. In an interview conducted 20 years later, he had this to say about his resignation: 'I went to SA Breweries in '62 for various reasons – don't want to talk about that. I fell out with the establishment, not the CSIR, but the Government'[36(p.2)].

Conclusion

Although this is one relatively small-scale study conducted by social scientists in South Africa in the 1950s, it nevertheless was worthwhile to explore its wider context and implications.

Posel characterised apartheid as a project of white supremacy and of Afrikaner nationalism.[37(p.327)] This neatly captures a major contextual factor to explain why the project and thesis caused such trouble, namely that it contradicted quite fundamental aspects of apartheid's grand design. One such contradiction involved the policy that Africans did not belong as permanent residents in the urban areas of South Africa. This policy was largely administered by the DNA, from where the key participants in the Sherwood study were drawn. The research directly contradicted government policy, and in no uncertain terms, as Eiselen's second reading of its possible impact reflected.

An emphasis on science and being scientific is noticeable almost everywhere in the documents. Kingwill, in a history of the CSIR, commented that the 'the institute (the NIPR – JL) had played a pioneering role in the introduction of the scientific approach to personnel management practice in southern Africa'[38(p.268)]. Sherwood commented in her interview on the importance of her work to be regarded as scientific. The focus on science is not altogether surprising, as 'being scientific' was a typically modern approach to a problem, which provided information for informed and rational planning. Earlier I quoted the work of Posel[8], who demonstrated the importance of counting and classifying the population in the apartheid state.

Sherwood's research, however, shows that governing with science and numbers contains significant risks for a political party with an ideology like the National Party. The open-endedness of science contains an inherent threat that empirical data may contradict the ideological grounding of policies. In Posel's view,

> *We presume that scientific knowledge has served as one more powerful tool in the hands of the powerful – but the relation between scientific knowledge and political and economic power is more complex than that. The exercise of power was always contested, in many ways apartheid was a 'demented sort of rationalism'.*[8(p.138)]

In terms of psychology, the NIPR's work in the 1950s and 1960s is often characterised as based on Elton Mayo's human relations approach. Fullagar[39] claimed that the human relations focus in South African personnel psychology has often been described as having 'a cultural bias'. In other words, it 'has always tended to emphasize *cultural* rather than individual differences' (emphasis in original) because 'South African society encourages industrial psychologists to construct theories of black and white behaviour which ignore socio-political determinants'[39(p.15)]. Sherwood's study shows that such a criticism requires more nuance, and that ignoring political and socio-economic factors was not a *sine qua non* of the NIPR's work. Certainly, the work done for her thesis cannot be accused of viewing 'the organization from a closed-system perspective'[39(p.22)]. It was explicitly not reductionistic; instead, it interpreted the troubles experienced by the African clerks within the full range of socio-cultural and political factors that determined their performance and efficiency: job reservation, apartheid legislation, unequal distribution of opportunities, etc. Indeed, Coupe stated that 'not all psychological research in South Africa is narrowly subservient to the interests of the dominant class, and work is occasionally published which is subversive of the ruling ideology'[34(p.43)]. This is perhaps the most significant conclusion one could reach from these events, as made abundantly clear by the consequences for Sherwood's academic project.

Acknowledgements

This work is based on research supported in part by the National Research Foundation of South Africa (grant number 119117). I acknowledge the assistance of Adrian Brock of the University of Cape Town; Carol Crosley, the Registrar of the University of the Witwatersrand; and Nick and John Sherwood, sons of Dr Rae Sherwood.

Competing interests

I have no competing interests to declare.

References

1. Sherwood R. The African civil servant – A socio-psychological study [unpublished PhD thesis]. Johannesburg: University of the Witwatersrand; 1960.
2. Louw J. South Africa. In: Baker D, editor. The Oxford handbook of the history of psychology. Oxford: Oxford University Press; 2012. p. 43–59.
3. Sherwood R. The psychodynamics of race: Vicious & benign spirals. Brighton: Harvester Press; 1980.
4. Luruli NM, Mouton J. The early history of research funding in South Africa: From the Research Grant Board to the FRD. S Afr J Sci. 2016;112(5/6), Art. #2015-0097, 6 pages. http://dx.doi.org/10.17159/sajs.2016/20150097
5. National Archives of South Africa (NASA). Navorsing met betrekking tot die keuring van die Bantoestaatsamptenaar. Geskiedenis van die projek. [Research regarding the selection of the Bantu State official. History of the project.] NSRN Reference N/R/6/394 Part 3. Pretoria: NASA; 1960–1966. Afrikaans.
6. National Archives of South Africa (NASA). Untitled document from the DNA, 1954 April 13. Reference N/R/6/394 Part 1. Pretoria: NASA; 1954–1955.
7. Bonner P, Delius P, Posel D. The shaping of apartheid. Contradiction, continuity and popular struggle. In: Bonner P, Delius P, Posel D, editors. Apartheid's genesis, 1935-1962. Johannesburg: University of the Witwatersrand Press; 1993. p. 1–41.
8. Posel D. A mania for measurement: Statistics and statecraft in the transition to apartheid. In: Dubow S, editor. Science and society in southern Africa. Manchester: Manchester University Press; 2000. p. 116–142.
9. National Archives of South Africa (NASA). A study of the Bantu Civil Servant – Report on progress for the year October 1956 – October 1957. Reference N/R/6/394/1 Part 1. Pretoria: NASA; 1954–1955.
10. National Archives of South Africa (NASA). Progress report to the Supervising Committee, October 1956 to October 1957. Reference N/R/6/394 Part 3. Pretoria: NASA; 1960–1966.
11. Sherwood R. The Bantu clerk: A study of role expectations. J Soc Psychol. 1958;47:285–316. https://doi.org/10.1080/00224545.1958.9919247
12. Sherwood R. Motivational analysis: A comparison of job attitudes among African and American professional and clerical workers. Proc S Afr Psychol Assoc. 1956–7;7&8:27–28.
13. National Archives of South Africa (NASA). Notes from the Supervising Committee meeting, September 1955. Reference N/R/6/394 Part 3. Pretoria: NASA; 1960–1966.
14. Minutes of a meeting of an ad hoc committee, dated 9 October 1973. University of the Witwatersrand, Faculty of Arts, Central Records Office, Archives Section, Document S73/811 (F.A. 102); 1973.
15. National Archives of South Africa (NASA). Letter of 3 October 1960 to the Registrar of the University of the Witwatersrand, Johannesburg. Reference N/R/6/394 Part 3. Pretoria: NASA; 1960–1966.
16. National Archives of South Africa (NASA). Letter of 14 November 1960 to the Supervising Committee. Reference N/R/6/394 Part 3. Pretoria: NASA; 1960–1966.

17. National Archives of South Africa (NASA). Letter of 7 November 1960 to P.M. Robbertse. Reference N/R/6/394 Part 3. Pretoria: NASA; 1960–1966.

18. National Archives of South Africa (NASA). Letter of 7 December 1960 to P.M. Robbertse. Reference N/R/6/394 Part 3. Pretoria: NASA; 1960–1966.

19. National Archives of South Africa (NASA). Letter of 31 July 1961 to the Secretary of DEAS. Reference N/R/6/394 Part 3. Pretoria: NASA; 1960–1966.

20. National Archives of South Africa (NASA). Letter of 14 September 1961 to the Registrar, University of the Witwatersrand. Reference N/R/6/394 Part 3. Pretoria: NASA; 1960–1966.

21. National Archives of South Africa (NASA). Letter of 7 September 1961 to the Secretary of DEAS. Reference N/R/6/394 Part 3. Pretoria: NASA; 1960–1966.

22. National Archives of South Africa (NASA). Letter of 15 June 1962 to the Secretary of DEAS. Reference N/R/6/394 Part 3. Pretoria: NASA; 1960–1966.

23. National Archives of South Africa (NASA). Letter of 1 August 1962 to the Secretary of DEAS. Reference N/R/6/394 Part 3. Pretoria: NASA; 1960–1966.

24. National Archives of South Africa (NASA). Letter of 27 February 1962 to Rae Sherwood. Reference N/R/6/394 Part 3. Pretoria: NASA; 1960–1966.

25. National Archives of South Africa (NASA). Letter of 2 July 1961 to the President of the CSIR. Reference N/R/6/394 Part 3. Pretoria: NASA; 1960–1966.

26. National Archives of South Africa (NASA). The selection and work efficiency of Bantu public servants. Interim report, National Institute for Personnel Research, December 1961. Reference N/R/6/394 Part 3. Pretoria: NASA; 1960–1966.

27. National Archives of South Africa (NASA). Letter of 25 September 1962 to the Secretary of the DEAS. Reference N/R/6/394 Part 3. Pretoria: NASA; 1960–1966.

28. National Archives of South Africa (NASA). Letter of 6 October 1965 to the Secretary of the DEAS. Reference N/R/6/394 Part 3. Pretoria: NASA; 1960–1966.

29. National Archives of South Africa (NASA). Letter of 20 October 1965 to W.M.M. Eiselen. Reference N/R/6/394 Part 3. Pretoria: NASA; 1960–1966.

30. National Archives of South Africa (NASA). Letter of 28 October 1965 to P.M. Robbertse. Reference N/R/6/394 Part 3. Pretoria: NASA; 1960–1966.

31. National Archives of South Africa (NASA). Letter of 28 February 1996 to the Secretary of the DEAS. Reference N/R/6/394 Part 3. Pretoria: NASA; 1960–1966.

32. National Archives of South Africa (NASA). Letter of 4 March 1968 to the Registrar, University of the Witwatersrand. Reference N/R/6/394 Part 3. Pretoria: NASA; 1960–1966.

33. Mann JW. Report on the revised thesis entitled: The African civil servant – A socio-psychological study submitted by Mrs R. Sherwood for the degree of Doctor of Philosophy. University of the Witwatersrand, Faculty of Arts, Central Records Office, Archives Section, Document S73/810 (F.A. 101): 1973.

34. Coupe S. Testing for aptitude and motivation in South African industry: The work of the National Institute for Personnel Research, 1946-1973. Bus Hist Rev. 1996;70:43–68. https://doi.org/10.2307/3117219

35. Biesheuvel S. Sherrington Club presentation on 12/8/88: Reminiscences about IQ testing and human groups. Handwritten document in Biesheuvel Archives, University of the Witwatersrand, Johannesburg, Box AU8 BIE; 1988.

36. Biesheuvel S. Interview with I. van W. Raubenheimer. Transcript available in the Raubenheimer Collection, Special Collections Section, University of Stellenbosch Library, Box 399 'Industriële sielkundiges'; 1982.

37. Posel D. The apartheid project, 1948-1970. In: Ross R, Kelk Mager A, Nasson B, editors. The Cambridge history of South Africa. Vol. 2, 1885-1994. Cambridge: Cambridge University Press; 2012. p. 319–368.

38. Kingwill DG. The CSIR – The first 40 years. Pretoria: CSIR; 1990.

39. Fullagar C. Organizational behaviour in South Africa: An historical overview. In: Barling J, editor. Behaviour in organizations. Johannesburg: McGraw-Hill; 1983. p. 3–26.