South African Journal of Science

volume 116
number 9/10

Trusting in peer review

Legal-ethical guidelines for human germline editing

Contributions of ruminant production to GHGs and human nutrition

Newly identified hominin trackways on Cape south coast

A classification algorithm for electricity fraud detection

ASSAf
ACADEMY OF SCIENCE OF SOUTH AFRICA

Volume 116
Number 9/10
September/October 2020

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: 051 444 2552
E: publish@sunbloem.co.za

Correspondence and enquiries
sajs@assaf.org.za

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

Research Article

Research Letter

Cover caption
Ruminant production plays a vital role in human nutrition and livelihoods. In an article on page 115, Scholtz and colleagues provide a balanced perspective on this role and the contribution of ruminants to greenhouse gas emissions.

Peer review: Not in crisis … but vibrant in its growing diversity

Not in crisis…but vibrant in its growing diversity.[1(p.49)]

In this issue of the *South African Journal of Science* (SAJS), we commemorate international Peer Review Week (21–25 September 2020). This annual event began informally in September 2015 as a partnership between ORCID®, the Open Researcher and Contributor ID, and the American Association for the Advancement of Science, the publisher of the *Science* journals. They initiated a powerful international conversation encouraging scholars to reflect regularly on the multiple features of peer review, to re-assess best practice, to debate strengths and weaknesses, and to examine possible enhancements.

With increased support the following year it was decided to have an annual theme which, in 2016, was 'Recognising peer review'. The topic was thus broadened from publications to include other areas of academic peer evaluation, like grant and employment applications and conference abstracts. In 2017, the week was linked to an international congress on 'Transparency in peer review', while 2018 concentrated on 'Diversity in peer review', considering situations beyond the Global North to include racial and gender diversity. Apparently, this topic was only 'Somewhat successful … [and] … remains a challenge'.[2] Last year, when 29 organisations were involved, the theme was 'Quality in peer review'.[3]

The year 2020 has been shattering, with the COVID-19 pandemic throwing societies and economies, along with the lives of researchers and the work of scholarly publishing, into disarray. As we confront fake news, a social media frenzy, scientific uncertainty and a crisis in academe, the chosen theme for 2020 is appropriately 'Trust in peer review'.[4]

The history of peer reviewing in South Africa has been neglected. The SAJS began more than a century ago as the *Proceedings of the South African Association for the Advancement of Science* and robust discussions after presentations at annual meetings were certainly a form of peer appraisal, but there is no clear evidence of when external peer review before publication became the Journal's convention.

Many academics suppose that external peer reviewing is well entrenched and has existed in its current form for an extremely long time, but this is not so and the analyses of Newman[5] and Baldwin[6] are informative in this regard. The process was haphazard and editorial judgement was the norm before peer review established itself in the late 20th century, and it evolved only because science funders (generally governments) wanted assurance from more than the researchers themselves that a funding investment would be scientifically rewarding. *Nature* began external peer review only in 1973 and the majority of science journals did so during the 1970s and 1980s; peer review accelerated as 'publish or perish' became endemic.

In this issue we present some current concerns around trust and peer review. Wolfgang Preiser and Rika Preiser grapple with the underlying concept of trust, arguing that COVID-19 may be re-shaping how scientific knowledge is verified. Growing numbers of retracted articles suggest many shortfalls in peer review and results communicated in preprints and the media may be, if not actually incorrect, certainly less than thorough. The effect of distrust of the scientific process itself is worrying, but the authors conclude that exhaustive peer review is – more than ever – absolutely vital to defend reliable science.

As Editor of the *South African Medical Journal*, Bridget Farham is in the vortex of COVID-19 problems relating to preprints and pressure on journals to fast track manuscripts. She provides concrete examples of how peer review is the process that most ensures trust, acknowledging that – whatever its flaws – it remains the gold standard for publishing reliable research results.

Keyan Tomaselli is also concerned by the numerous retractions of premature research that erode trust. In his words: 'This is not necessarily fake science, but potentially good science managed badly.' However, his view is that oversight through South African institutions, such as the Academy of Science of South Africa (ASSAf), the National Research Foundation (NRF) and the Department of Higher Education and Training, ensure high quality by supporting trustworthy peer review.

In his contribution, Robin Crewe, the Chair of ASSAf's Committee on Scholarly Publishing in South Africa, explains how ASSAf assures the quality of South African journals as a whole. Analogous to the peer review undertaken *by* journals of research manuscripts submitted to them, the external peer review of groups of journals is entirely novel. In this way, trust in the journals themselves is secured by review panels that identify flaws and recommend improvements. These reviews add a further element of trust in the work of South African academics who publish in these journals. ASSAf is also responsible for producing the 'Code of Best Practice in Scholarly Journal Publishing, Editing and Peer Review'[7], a further safeguard against malpractice.

ASSAf's role is also highlighted by Johann Tempelhoff who reports on a peer review webinar hosted by the Academy on 31 July 2020 that attracted more than 300 registrants and generated considerable discussion. Chaired by Lucienne Abrahams, the presenters included Salmina Mokgehle, an Associate Editor Mentee of SAJS. Discussion was lively and critical, but as Tempelhoff concludes, the Pandora's box of peer reviewing brings forth more positive than negative results for knowledge inquiry, which is not to say that improvements cannot be made.

Steven Johnson follows up on an earlier publication[8] in which he scrutinised the role of the NRF in the peer-review process. Given that the NRF's academic rating process engages peer reviewers in assessing researchers' CVs, while the *h*-index is a measure of the productivity and citation impact of the publications of researchers, Johnson investigated whether the two were satisfactorily aligned for researchers in the biological sciences. He concludes that this is the case, lending weight to trusting in good peer review at all levels of scholarship.

Academics and publishers must ensure that trust is not eroded and that the process is equitable and fair to everyone involved in order to produce reliable evidence-based knowledge.

References

1. Ware M. Peer review: Recent experience and future directions. New Rev Inf Netw. 2011;16(1):23–53. https://doi.org/10.1080/13614576.2011.566812
2. Global state of peer review 2018. Publons/Clarivate Analytics; 2018. https://doi.org/10.14322/publons.GSPR2018
3. Meadows A. Peer review week is five! The Scholarly Kitchen. 2019 June 26 [cited 2020 Aug 31]. Available from: https://scholarlykitchen.sspnet.org/2019/06/26/peer-review-week-is-five/
4. Meadows A. In peer review week we trust. The Scholarly Kitchen. 2020 April 09 [cited 2020 Aug 31]. Available from: https://scholarlykitchen.sspnet.org/2020/04/09/in-peer-review-week-we-trust/
5. Newman B. Authorising geographical knowledge: The development of peer review in *The Journal of the Royal Geographical Society*, 1830–c.1880. J Hist Geogr. 2019;64:85–97. https://doi.org/10.1016/j.jhg.2019.03.006
6. Baldwin M. Scientific autonomy, public accountability, and the rise of 'Peer Review' in the Cold War United States. Isis. 2018;109(3):538–558. https://doi.org/10.1086/700070
7. Code of Best Practice in Scholarly Journal Publishing, Editing and Peer Review [webpage on the Internet]. c2018 [cited 2020 Aug 17]. Available from: https://sites.google.com/view/assaf-nsef-best-practice
8. Lovegrove BG, Johnson SD. Assessment of research performance in biology: How well do peer review and bibliometry correlate? Bioscience. 2008;58(2):160–164. https://doi.org/10.1641/B580210

HOW TO CITE:
Carruthers J. Peer review: Not in crisis … but vibrant in its growing diversity. S Afr J Sci. 2020;116(9/10), Art. #8853, 1 page. https://doi.org/10.17159/sajs.2020/8853

AUTHOR:
Robert Morrell[1]

AFFILIATION:
[1]Programme for the Enhancement of Research Capacity (PERC), School of Education, University of Cape Town, Cape Town, South Africa

CORRESPONDENCE TO:
Robert Morrell

EMAIL:
robert.morrell@uct.ac.za

HOW TO CITE:
Morrel R. Bill Freund (1944–2020): Economic historian. S Afr J Sci. 2020;116(9/10), Art. #8827, 1 page. https://doi.org/10.17159/sajs.2020/8827

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

PUBLISHED:
29 September 2020

Bill Freund (1944–2020): Economic historian

Bill (William Mark) Freund died suddenly in Durban in the early hours of Monday, 17 August 2020. The virtual memorial event held the following Saturday and attended by over 170 people testified to his scholarship, his intellectual influence, and to the vast network of friends and admirers that he had built up in a career that began with the completion of a Yale PhD in 1971.

Bill was a major figure in the study of South African history. He published six single-authored books, the first in 1981, *Capital and Labour in the Nigerian Tin Mines* (Longman), and the latest in 2018, *Twentieth Century South Africa: A Developmental History* (Cambridge University Press). In between these he wrote numerous articles, many chapters in books and reviews, co-edited two books and graduated 12 PhD students.

Bill was born on 6 July 1944 in Chicago (Illinois, USA), the only child of Austrian émigrés Carlo Freund (1902–1989) and Elisabeth Gross (1912–1995). His parents (not yet married) managed to escape the Holocaust, leaving from Trieste by ship bound for the USA in the second half of 1939. Most of their relatives who remained behind in Europe perished before the war ended in 1945.

Bill's Yale PhD was his entrée into the study of Africa. In his work on the Batavian period at the Cape (1803–1806) he displayed aspects of his work that were to become his signature over the coming decades. These were a facility and willingness to work in languages other than English, his ability to absorb enormous amounts of information and rapidly to make sense of it, his confidence in merging detail with 'the big picture', and his passion for writing about and understanding Africa.

Before arriving in 1985 as Professor of Economic History at the (then) University of Natal in Durban, Bill had already met with institutional adversity. Despite spells of employment at Yale, Harvard, and Kirkland College, New York, he was never able to secure a tenured job. On the other hand, he had happy work experiences at the University of Dar-es-Salaam and Ahmadu Bello (Zaria, Nigeria) as well as a period at the African Studies Institute at Wits University with Charles van Onselen.

He got his big break with the job in Durban. It was here that he discovered security, affirmation, and happiness. He found pleasure in becoming a regular part of a weekly touch rugby game, gaining an acceptance that had often eluded him in the country of his birth.[1] In Durban, his prodigious gifts of scholarship were expressed to the full. Bill arrived in Durban on the back of the publication of his second book, *The Making of Contemporary Africa* (Indiana University Press; 1984). As John Lonsdale noted in his review in the *Journal of African History*, this was 'a landmark in African historiography'[2(p.122)]. The influence of this book can be gauged by its having been revised twice, the third edition coming out in 2016.

Some of the strongest academic ties that Bill created were with colleagues in Development Studies. In 1986, Bill, Mike Morris and Gerhard Maré began the journal *Transformation*, which recently published its 100th issue. The journal was characterised by political engagement, critique and independence. The editorial team was joined shortly after the start by Vishnu Padayachee, at that time working at the University of Durban-Westville. Bill continued to work closely with Vishnu, being drawn into studies of Durban's Indian population with *Insiders and Outsider: The Indian Working Class of Durban in the Twentieth Century* (Pearson Education; 1995), as well as into a close socio-political and economic analysis of Durban, the city. He and Vishnu co-edited *The D(urban) Vortex: A South African City in Transition* (University of KwaZulu-Natal Press; 2002).

Bill was primarily an academic and activism did not come naturally to him. Nevertheless, as the end of apartheid loomed, his expertise in political economy was recognised and he became part of the Economic Trends Group, in turn part of the Macro-Economic Research Group set up by the ANC in 1991 to develop economic policy. Economic Trends was a COSATU-inspired sub-group in which fellow Durbanites Alec Erwin and Stephen Gelb were important figures.[3]

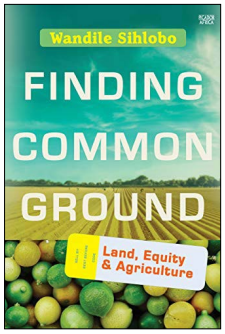
But Bill will be remembered primarily as an economic historian and a materialist. He was concerned with capital accumulation and labour relations, and his particular interest in, and sympathy with, workers is evident in the vast corpus of his work. Perhaps the best example is the magisterial condensation of worker history in Africa simply titled *The African Worker* (Cambridge University Press; 1988). Although not always easy to discern in his work, he was moved by the exploitation of workers and often outraged by the injustices of the workplace.

Bill's work illuminated realities at the local, national, regional and global levels. He connected people to one another, built ideas about Africa, and, through his writings, broadcast these to the world. His autobiography, *Bill Freund: An Historian's Passage to Africa,* will appear in 2021.

References

1. Morrell R. Touch rugby, masculinity and progressive politics in Durban, South Africa, 1985-1990. Int J History Sport. 2017;34(7–8):619–638. https://doi.org/10.1080/09523367.2017.1379996
2. Lonsdale J. A materialist history of Africa: *The making of contemporary Africa: The development of African society since 1800*. By Bill Freund. London: Macmillan; 1984. J Afr History. 1985;26(1):122–123. https://doi.org/10.1017/S0021853700023185
3. Freund W. Swimming against the tide: The Macro-Economic Research Group in the South African transition 1991–94. Rev Afr Political Econ. 2013;40(138):519–536. https://doi.org/10.1080/03056244.2013.854038

BOOK TITLE:
Finding common ground: Land, equity and agriculture



AUTHOR:
Wandile Sihlobo

ISBN:
9781770107168 (softcover, 239 pp)

PUBLISHER:
Picador Africa, Johannesburg; ZAR299

PUBLISHED:
2020

REVIEWER:
William Beinart[1]

AFFILIATION:
[1]African Studies Centre, University of Oxford, Oxford, United Kingdom

EMAIL:
william.beinart@sant.ox.ac.uk

HOW TO CITE:
Beinart W. Strategies for land reform and agriculture. S Afr J Sci. 2020;116(9/10), Art. #8363, 1 page. https://doi.org/10.17159/sajs.2020/8363

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

PUBLISHED:
29 September 2020

Strategies for land reform and agriculture

Wandile Sihlobo has established himself over the last 5 years as one of the leading journalists and commentators covering land and agricultural issues in South Africa. He writes for *Business Day*, *Daily Maverick* and *Farmers Weekly* and this book is – in his words – a 'harvest' or 'isivuno' of his articles with linking passages. Since 2016 he has worked at AgBiz – an association of agribusiness organisations that aims to publish data and influence policy. He has been incorporated into key policy groups such as the President's Advisory Panel on Land Reform and Agriculture (PAP 2019). Trained as an agricultural economist, Sihlobo has wonderful access to information on agriculture and land, to both government and private sector, and to breaking issues. He gets out and about, visits farms and projects, talks to a wide range of practitioners, and attends conferences. Above all, *Finding Common Ground* is a well-informed, broad-ranging and sensible book that covers a wide variety of topics from land reform to international markets. Some articles are jointly written, and it is not always clear when we are hearing Sihlobo himself. But collectively these pieces show a deep understanding of South African agriculture – its diversity, strengths and weaknesses – in comparative perspective. He is also strongly aware of the politics of land, although he approaches the politics with caution.

Sihlobo is a moderniser and argues that land reform should be focused not least on production. He and his co-authors, especially Johann Kirsten in a 2018 series, emphasise that successful agriculture at any scale requires investment, intensification, skills and technology. They note the expansion of high-value fruit, nuts and horticulture from old-established citrus to more novel blueberries. In 2018, nearly 50% of the value of agricultural output was exported – much of it from this sub-sector, which is also important for internal markets and food security. Sihlobo is an advocate of GM seeds and sees them as fundamental to South Africa's continued capacity in maize, where yields have increased recently; he argues that such innovations should be extended to smallholders. There is an intriguing article analysing regional approaches to the legalisation of dagga; he supports this as a means of stimulating smallholder production and high-value processed products.

Sihlobo and Kirsten are particularly focused on how to pursue land reform in a way that may sustain and enhance production, livelihoods and employment. They discuss thoroughly the problem of who should be beneficiaries and what kinds of transfers facilitate farming. For them, expropriation without compensation will have a negative impact, on both agricultural investment and financial institutions. In these and other articles, Sihlobo accepts that agriculture and associated rural livelihoods will be pursued at all scales. He returns on a few occasions to the potential of perhaps 1 million hectares of underutilised land in the former homelands, which could be a significant source of income for the relatively poor communities that live there. He sees particular promise in partnerships where commercial operations or experts work with smallholders and new black farmers – providing knowledge, inputs, capital and access to markets: he is impressed by a visit to a 65-ha community-owned farm on which blueberries and peppers are grown, and sounds a particular note of optimism about a group of 17 black commercial farmers in Matatiele, initially supported by the Old Mutual Masisizane fund, that is expanding maize and wheat production.

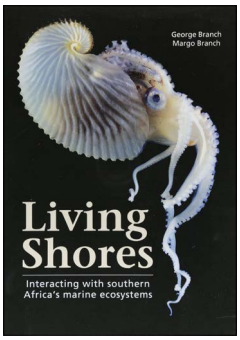
The problem with land reform in this view is not an insufficiency of land but ineffective bureaucracies and inadequate support to beneficiaries. Too many have been 'set up for failure'; government departments have been slow to provide opportunities and back-up. Sihlobo hopes that the recent re-amalgamation of the Departments of Agriculture and Land Reform/Rural Development, under Thoko Didiza, will resolve some of the blockages. He and Kirsten are uncomfortable with the government's shift, under the Proactive Land Acquisition Strategy (PLAS, from 2006), to renting out redistributed land. While they are not in favour of rapid privatisation in the former homelands, they see individual land ownership as the most likely form of tenure to facilitate agricultural investment and production.

The joint position paper that Sihlobo wrote with three other members of the PAP suggests that individual smallholder models are the most successful. By contrast, South Africa has tended to pursue group farming models through Communal Property Associations where decision-making can be a recurring hurdle. It is certainly interesting to see two papers that fed into the PAP and this one also, somewhat surprisingly, seems to favour a fast-track, state-supported programme to transfer 30% of agricultural land in freehold to black South Africans. It is not exactly clear what the authors mean, and whether this includes the 10% of agricultural land that has been transferred through government schemes. But this proposal seems somewhat out of step with Sihlobo's approach as a whole, in which he puts projects and successful production ahead of unrealistic targets. Elsewhere, he is highly critical of Zimbabwe's fast track.

Although he does not develop this point, Sihlobo also notes that downstream processing of agricultural produce generates a greater percentage of GDP than primary agricultural production and such industries and processes have particular potential for jobs, especially those that are highly skilled. Is the implication that unravelling the large farm sector, for an uncertain future in which production and food security is dependent on undercapitalised smallholders, is unwise and that a more productive priority is deracialising the 'entire production chain'? It would be valuable to have further development of Sihlobo's views on some of these ideas, as well as more detailed discussion about interventions that might unlock the productive potential of smallholders.

This book is a very welcome intervention – an optimistic discussion with a strong sense that, despite the legacies of apartheid and divisive politics, problems can be solved. It is the approach of a well-informed, sensitive observer who advocates building on successful models of production and technical advances. This is what South Africa needs.

BOOK TITLE:
Living shores: Interacting with southern Africa's marine ecosystems



AUTHORS:
George Branch and Margo Branch

ISBN:
9781431700813 (hardcover, 336 pp)

PUBLISHER:
Penguin Random House, Cape Town; ZAR490

PUBLISHED:
2018

REVIEWER:
Judy Mann-Lang[1]

AFFILIATION:
[1]South African Association for Marine Biological Research, Durban, South Africa

EMAIL:
jmann@saambr.org.za

HOW TO CITE:
Mann-Lang J. Celebrating a century of marine discoveries in South Africa. S Afr J Sci. 2020;116(9/10), Art. #8378, 2 pages. https://doi.org/10.17159/sajs.2020/8378

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

PUBLISHED:
29 September 2020

Celebrating a century of marine discoveries in South Africa

When the first edition of *The Living Shores of Southern Africa* came out in 1981 it very quickly became a firm favourite of professional scientists, a favourite textbook used by lecturers and students at universities around South Africa, a favourite book for amateur marine citizen scientists around the coast and, perhaps most importantly, a favourite book for hundreds of budding marine biologists, young people fascinated by the ocean, for whom the book revealed a new world. What made this book stand out was the way that it was written – in a language accessible to all – and illustrated, in clear black and white drawings and excellent photographs. The book invited the reader to share the authors' fascination with our living shores and in so doing inspired the current generation of South African marine scientists and educators.

There was a great deal of excitement when George and Margo Branch made it known that they were working on a new edition (after nine reprints the 1981 edition had become dated). And, after many years of eager anticipation, the *Living Shores: Interacting with Southern Africa's Marine Ecosystems* was published in 2018. This book is not just an updated version of the popular 1981 *Living Shores*, but a whole new book. And when you start to read it, the reason it took so long to write becomes clear.

The selection of topics must have been incredibly difficult, but the authors have succeeded in including a broad range and have managed to cover each topic in sufficient depth to make each chapter a meaningful overview. The book is divided into two parts – 'The Ecosystems' and 'The Human Factor', each sub-divided into several chapters. Part one, 'The Ecosystems', consists of nine chapters – starting with an introduction to the world's ocean in the 'Surging seas'. From the properties of sea water to the global currents, seabed mapping, waves and tides, the chapter then focuses attention on southern Africa, its currents, coastline, biodiversity and climate, thereby providing a sound introduction to the rest of the book. The chapter on 'Restless continents' focuses on the geology underlying the coastline and gives a solid foundation upon which the other chapters are built. Naturally, given George and Margo's passion, the following chapter is the longest in the book and explores the 'Rocky shores'. It is in this chapter that the depth and breadth of science covered in the book becomes clear. In this one chapter the authors have managed to convey the results of hundreds of papers in a way that draws the reader in, eager to learn, as they explore this ecosystem through the many stories of complex interactions between animals and between animals and plants, and fascinating adaptations of species to their harsh environment. A myriad of biological mysteries is resolved on this journey through South African researchers' work on this ecosystem. The chapter on 'Sandy beaches' is notable for its ability to transform this often-overlooked ecosystem into a treasure trove of fascinating stories about beaches, their formation, and the life that they support. The vital role of estuaries to humans is clearly addressed in the 'Estuaries' chapter, which also contains recent information about the variable nature of estuaries and the importance of appropriate management of these complex ecosystems. Most people are unaware of the fact that the South African coastline is dotted with 30 rocky outcrops large enough to qualify as islands. In the chapter covering 'Islands', the fascinating life on these offshore rocky outcrops is revealed. Rather than being offshore refuges, safe from human impact, this chapter describes the plunder that has taken place over centuries. From seals to seabirds, guano to eggs, nothing has been safe from human impact. Back to the coast, 'Kelp forests' is a richly illustrated chapter filled with new insights and fascinating stories of the interactions between the various components of this ecosystem. The implications of these complex interactions for the management of some of South Africa's most valuable marine resources – abalone and rock lobster – are also addressed. The diversity and complexity of South Africa's 'Coral reefs' and the difficult to study connections between species within 'The open ocean' bring part one to a close.

While much of the science that has been undertaken in South African coastal ecosystems may be thought of as 'fundamental' science, the complex interactions between species that has been revealed has critical implications for the management of many valuable marine resources. And this leads us to the next part of the book. 'The Human Factor' starts with a look backwards. 'The Cradle of Mankind' introduces the reader to our distant past and explains the links between the evolution of humans and coastal resources. Moving rapidly into the present and with a few glimpses into the future, the following nine chapters deal with human use, and abuse, of the ocean. Given human reliance on seafood, the 'Harvest of the seas' chapter is a natural starting point. This chapter provides a succinct introduction to the different resources harvested along the South African coast. The sobering figures in the chapter clearly demonstrate how, with some notable exceptions, many of South Africa's marine resources have been overexploited. The next chapter, 'Sustaining the catch', provides an excellent overview of marine resource management in South Africa. Complicated topics such as stock assessments, Maximum Sustainable Yield, bioeconomic models, operational management plans, ecological risk assessments, the ecosystem effects of fishing' and the even more difficult issues of co-management are explained clearly, without losing detail. The next few chapters introduce issues facing our coast that are not as well known, but equally important. In 'March of the aliens' the impact of alien marine plants and animals on biodiversity is covered – including their origins and possible management solutions. While it is often difficult to understand the impact of 'Development & Pollution' on the ocean, the impact on the coast is more obvious. These impacts are clearly demonstrated through well-illustrated and relevant case studies, while the impacts of a wide variety of sources of pollution are also addressed. 'Marine mining' has recently become topical and this chapter provides a good introduction to the origins, value and impact of this industry. 'Climate change' is one of those topics that many people talk about, but few truly understand. In this chapter, climate change is explained clearly and the complex interplay between the ocean and the climate is addressed. In fact, this chapter, with its clear artwork and photos could be used by anyone as a fundamental introduction to climate change, and its impact.

The last three chapters, 'Spreading sea wisdom', 'Marine protection' and 'Policy', ensure that the book ends with a look into the future. In 'Spreading sea wisdom', the crucial role of marine education, and of inspiring care for the ocean is discussed. Given South Africa's track record in Marine Protected Areas (MPAs), it is fitting that this tool in the conservation toolbox should be included. 'Marine protection' covers everything from the rationale for MPAs to the selection of where and how large MPAs should be, assessments of the effectiveness of MPAs and, critically, MPA management challenges and solutions. The last chapter on 'Policy' pulls together the previous chapters on human interaction with the ocean and provides an overview of South Africa's plethora of policies pertaining to the ocean. While the implementation of these policies is often found wanting, the spirit with which they were drafted remains a powerful message of hope.

Living Shores is so many things.

It is an excellent **reference** book. As described above, the science in the book represents the best of South Africa's marine science over the last 60 years. The sheer volume of research covered is astounding, as the book provides a synthesis of over 1250 papers, books and research reports. More remarkable is that the complexity of the research has been explained in such a way as to be not only understandable but also fascinating. It should be the recommended textbook for every undergraduate studying marine biology, every marine communicator, and anyone passionate about the ocean. There is no comparable text that I am aware of anywhere in the world.

Just as its predecessor did, this is a book that **inspires**. The book is richly illustrated to portray the diversity and magnificence of our living shores. The photographs (most from George and Margo's own collection), artwork, diagrams and maps deserve a special mention – they complement and bring the text alive and make many of the more complex research results and concepts much easier to understand. The book is filled with interesting anecdotes and stories that reinforce a personal connection with marine life.

The book is an example of **science communication** at its best. George and Margo Branch were science communicators long before it became fashionable. George's many years of lecturing and Margo's skills as an artist and experience in marine education are evident in the style of writing, the selection of stories and the illustrations. Good science communication relies increasingly on 'telling stories' and the book is filled with such stories – stories that are easy to share and that marine educators will be able to use, again and again.

Most importantly the book is a **celebration** – of our living shores, the landscapes, the plants, the animals and, most importantly, the people who have dedicated their lives to exploring and conserving it. Unusually in a book of this nature, the authors have elected to name individuals in the text. Rather than a list of impersonal references, they have introduced individuals, almost all of whom they know personally after their lifetime of work in marine research and conservation in South Africa. From Gilchrist Medal winners to young and upcoming students, the book reflects the life work of hundreds of passionate marine scientists, educators and conservationists. This feature of the book serves several purposes. Scientific writing is generally depersonalised, meaning that most people have little connection to science. People connect with people first, then science. Using names in the book reconnects people to the research and brings that critical human element into the research. The use of names is also a wonderful way to recognise the many people who have dedicated their lives to South Africa's marine research and conservation. And lastly, the names generate a sense of pride, pride in our colleagues and friends, and pride in South Africa's marine science.

This is a book to cherish, to enjoy, to explore and to share with the next generation of young marine champions. It is a book to inspire our youth to continue the work that George and Margo Branch have celebrated so beautifully.

Academic publishing in pandemic times

AUTHORS:
Wolfgang Preiser[1,2]
Rika Preiser[3]

AFFILIATIONS:
[1]Division of Medical Virology, Department of Pathology, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa
[2]National Health Laboratory Service, Cape Town, South Africa
[3]Centre for Complex Systems in Transition, Stellenbosch University, Stellenbosch, South Africa

CORRESPONDENCE TO:
Wolfgang Preiser

EMAIL:
preiser@sun.ac.za

HOW TO CITE:
Preiser W, Preiser R. Academic publishing in pandemic times. S Afr J Sci. 2020;116(9/10), Art. #8803, 2 pages. https://doi.org/10.17159/sajs.2020/8803

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
peer review, scientific discourse, COVID-19, public trust in science

PUBLISHED:
21 September 2020

Even though it tends to feel like ages, it has not been that long since the final days of 2019, when cases of severe respiratory illness (now known as COVID-19), caused by a previously unknown coronavirus (since named SARS-CoV-2), were reported from China. The ongoing COVID-19 pandemic has brought unprecedented disruption to almost every area of our daily and professional lives.

Science has not been spared, nor has scientific publishing. Most researchers have been unable to continue with their work, and many had to all but re-invent their teaching. Quite a few have re-invented themselves as coronavirus researchers.[1] As biomedical researchers, we are astonished to see how much interest the public is taking in our findings. For no other disease do members of the public so fervently seek out reports in traditional and social media about the latest research findings. These reports often trigger controversial discussions, mostly on social media platforms, about rather complicated aspects of epidemiology, diagnostics, pathogenesis or therapy. Many of these issues are matters outside the realm of everyday life and normally left to experts to assess the evidence and translate it into practice.

At the same time, public health policymakers urgently need scientific findings as a basis for measures and policies to control the pandemic or to mitigate its consequences. This instils an almost unprecedented sense of urgency for scientists to produce findings and results which is re-shaping the traditional ways in which scientific knowledge has previously been verified. Answers must be found and measures implemented in real time to respond to a rapidly unfolding situation for which no-one has a 'recipe book', as the world has never before faced a pandemic caused by an infectious agent with the characteristics of SARS-CoV-2.

Add to this the unfortunate politicisation of public health measures like the universal wearing of non-medical (cloth) masks, re-opening of schools etc., which in many places are used to support certain ideologies.

Science thus finds itself in a tricky situation. There is a great and urgent need for relevant studies, and there is a flood of funding opportunities which are very tempting, especially as other funding opportunities are expected to dwindle, and there is a significant increase in calls for papers and offers of expedited review. Researchers are heeding the call: as of today (17 August 2020), a simple PubMed search for the terms "COVID-19" OR "SARS-CoV-2" OR "2019-nCoV" yields an astonishing 40 660 results.

There are many pressures and incentives to try and be quick, even though that may be at the expense of thoroughness. It has become the norm to disseminate scientific data without, or prior to, peer review by means of preprint servers, news releases, news reports, and articles on science outreach platforms like *The Conversation*.

If done irresponsibly, this can cause harm – for example by touting unproven treatments with 're-purposed' drugs. One such drug, chloroquine, was said to be beneficial based on a small number of COVID-19 patients treated early on in the pandemic. Such observations should only serve as first leads but subsequently need to be put to the test by performing proper randomised controlled trials. In the meantime, this observational case series has been published, with critical reflection by the editors on weaknesses of the reported work and the merits of publishing data obtained in a sub-optimal manner.[2]

The avalanche of unscientific claims about the benefits of chloroquine sparked by French microbiologist Raoult soon reached the White House in Washington DC. A paper published by the prestigious journal *Lancet* purported to have conducted an analysis of a global hospital patient registry, finding that hydroxychloroquine was not only not beneficial but in fact linked to lower survival in hospitalised COVID-19 patients. After doubts were raised about the existence and quality of the underlying database, and after the registry was not made available for scrutiny, this paper and a second one in another top medical journal, claiming to use the same database for a different analysis, were retracted.[3,4]

A deluge of manuscripts related to the pandemic is being uploaded onto preprint servers; as of 17 August 2020, there are 6132 COVID-19 SARS-CoV-2 preprints on medRxiv and 1626 on bioRxiv. This is cause for concern, as these studies tend to be picked up by journalists before the 'safety net' meant to ensure the quality and integrity of science has fully unfolded.[5]

It is all the more worrying if papers are accepted and published by reputable journals despite serious shortcomings. The two vital components of scientific quality assurance – editorial oversight and peer review – are not foolproof, as the two retracted papers show. While it is comforting that the subsequent layers, critical reception by peers and replication by other studies, exposed the major flaws of these papers, the published retractions provide no evidence for critical reflection by the journal editors.[6] Such nonchalant retraction notices are not a new phenomenon; they might reflect a deeper problem, as the problem keeps recurring.[7] One might have wished that following the MMR vaccine–autism debacle, which continues to cause damage by reducing measles immunisation rates and favouring outbreaks, editors would have learnt their lesson.[8,9]

Lapses such as these are avoidable. Yes, the ongoing pandemic places an enormous burden on editors, reviewers and just about everyone else in the scientific community. Those best placed to provide meaningful peer review on submitted manuscripts are probably the same people who are themselves trying to obtain funding and ethics approvals, conducting trials, analysing data and writing manuscripts and thus do not have time to undertake peer reviews. On the other hand, 'informal' peer review may yield unexpected benefits, even before a paper is submitted to a journal. An example from the early days of the pandemic is the withdrawal of a preprint paper claiming that the SARS-CoV-2 genome contained elements from the genome of HIV.[10] Vigorous 'open' peer review,

taking place on science blogs and on Twitter, seems to have prompted the withdrawal of the manuscript.[5] This is to be welcomed.

A rational approach will go a long way, especially if supported by some knowledge of the field in question. For example, as most antiviral drugs are the end product of painstaking research conducted over decades, starting with basic virology and structural biology, then chemistry, then spending years and fortunes conducting clinical trials[11], it is highly unlikely that a decades-old antimalarial like chloroquine would have major, hitherto unrecognised, antiviral activity, just as it is unlikely that major toxicities not seen in decades would suddenly emerge. Both the poor design of an uncontrolled clinical study and the implausibility (it may well not exist) of an enormous clinical database underlying a major analysis could have been spotted by editors and reviewers alike. The urgency of coming up with therapeutic entities for COVID-19 patients may have clouded their judgement.

But is there a deeper problem? We would argue yes, definitely; and it goes beyond science. Scientists are familiar with the scientific process. Even though it may involve occasionally unrewarding experiences with editors and peer reviewers and sometimes unpleasant exchanges with colleagues, we by and large accept and engage in more or less constructive interactions which may or may not confirm results that themselves are the products of careful experimentation and analysis. It is its ultimately self-correcting nature that makes science robust.[12]

However, in the current pandemic climate, the eyes of policymakers, politicians and the public are on early results. Many are not familiar with how science works and may regard vigorous discussions among scientists, especially when earlier reports are proven wrong, as proof that science is unable to contribute meaningfully. Controversial discussions about the merits and shortcomings of scientific studies and the interpretation of research data are an indispensable component of science. These discussions usually happen 'behind the scenes' with little attention paid by anyone outside the field. Such a discourse can, however, be misinterpreted, or even abused, to paint a picture of discord and cluelessness. Fuelled by hyping of questionable data and reckless politicisation, this may undermine public trust in science and become a major problem, as has been evidenced with vaccines, climate change and the pandemic response.[13]

It may be challenging to communicate nuance, uncertainty and complexity to non-scientific audiences, but not doing so causes harm. Good, responsible reporting requires thorough reading of studies and speaking to experts, not just copying of press releases.[14] Simply contributing to the COVID-19 'infodemic' is not a solution.[15]

Academic institutions, publishers and editors must also assume responsibility. 'Speed science' must be discouraged, enticing as it may appear at first glance.[16] Poor-quality research, with studies that are biased or not properly controlled, that is statistically underpowered or simply repetitive, is a flash in the pan and does not advance science.[17] Journals should formulate and abide by clear policies on how to deal with the flood of COVID-19-related manuscript submissions in a responsible manner, and not by sacrificing quality for speed or volume (see Panda[18] for example).

As Phillips[19] wrote in a recent contribution to this journal, pandemics 'accelerate trends and tendencies already in train but not yet at full pace'. The COVID-19 pandemic seems to be having exactly these effects on scientific publishing. The pre-existing tendencies towards pre-publications being made public and towards open access publishing have gathered pace.[20]

Phillips[19] also mentions the propensity of pandemics to 'highlight the basic features of any society and its modus operandi, especially its shortcomings and fault lines, in ways which are difficult to ignore'. This, too, is very true for COVID-19. The pandemic has highlighted that 'bad science' happens when hype takes over or when people embark opportunistically on research outside their own specialist fields.

Fortunately, the scientific endeavour is by and large intact. Even during the pandemic it will call out poor research in the end. Yet to maintain or restore public trust in science, scholarly publishing needs to ensure high-quality peer review and be prepared to sacrifice high scores on the public relations side in the interest of maintaining the integrity of science.

References

1. Gibney E. The pandemic mixed up what scientists study – and some won't go back. Nature. 2020;582(7811):173–174. https://doi.org/10.1038/d41586-020-01525-z
2. Voss A, Coombs G, Unal S, Saginur R, Hsueh PR. Publishing in face of the COVID-19 pandemic. Int J Antimicrob Agents. 2020;56(1):106081. https://doi.org/10.1016/j.ijantimicag.2020.106081
3. Mehra MR, Ruschitzka F, Patel AN. Retraction: Hydroxychloroquine or chloroquine with or without a macrolide for treatment of COVID-19: A multinational registry analysis. Lancet. 2020;395(10240):1820. https://doi.org/10.1016/S0140-6736(20)31324-6
4. Mehra MR, Desai SS, Kuy S, Henry TD, Patel AN. Retraction: Cardiovascular disease, drug therapy, and mortality in Covid-19. N Engl J Med. 2020;382(26):2582. https://doi.org/10.1056/NEJMc2021225
5. Majumder MS, Mandl KD. Early in the epidemic: Impact of preprints on global discourse about COVID-19 transmissibility. Lancet Glob Health. 2020;8(5):e627–e630. https://doi.org/10.1016/S2214-109X(20)30113-3
6. Piller C, Servick K. Two elite medical journals retract coronavirus papers over data integrity questions. ScienceMag. 2020 June 04 [cited 2020 Aug 17]. Available from: https://www.sciencemag.org/news/2020/06/two-elite-medical-journals-retract-coronavirus-papers-over-data-integrity-questions
7. Collier R. Shedding light on retractions. CMAJ. 2011;183(7):E385–386. https://doi.org/10.1503/cmaj.109-3827
8. Fleck F. MMR controversy raises questions about publication ethics. Bull World Health Organ. 2004;82(4):311–312.
9. Greenhalgh T. Why did the Lancet take so long? BMJ. 2010;340:c644. https://doi.org/10.1136/bmj.c644
10. Pradhan P, Pandey AK, Mishra A, Gupta P, Tripathi PK, Menon MB, et al. Uncanny similarity of unique inserts in the 2019-nCoV spike protein to HIV-1 gp120 and Gag [preprint] (withdrawn). bioRxiv. 2020 January 31. https://doi.org/10.1101/2020.01.30.927871
11. De Clercq E, Li G. Approved antiviral drugs over the past 50 years. Clin Microbiol Rev. 2016;29(3):695–747. https://doi.org/10.1128/CMR.00102-15
12. Shamoo AE. Validate the integrity of research data on COVID 19. Account Res. 2020;27(6):325–326. https://doi.org/10.1080/08989621.2020.1787838
13. Sattui SE, Liew JW, Graef ER, Coler-Reilly A, Berenbaum F, Duarte-García A, et al. Swinging the pendulum: Lessons learned from public discourse concerning hydroxychloroquine and COVID-19. Expert Rev Clin Immunol. In press 2020. https://10.1080/1744666X.2020.1792778
14. Saitz R, Schwitzer G. Communicating science in the time of a pandemic. JAMA. 2020;324(5):443–444. https://10.1001/jama.2020.12535
15. The Lancet Infectious Diseases. The COVID-19 infodemic. Lancet Infect Dis. 2020;20(8):875. https://doi.org/10.1016/S1473-3099(20)30565-X
16. Dinis-Oliveira RJ. COVID-19 research: Pandemic versus "paperdemic", integrity, values and risks of the "speed science". Forensic Sci Res. 2020;5(2):174–187. https://doi.org/10.1080/20961790.2020.1767754
17. Glasziou PP, Sanders S, Hoffmann T. Waste in COVID-19 research. BMJ. 2020;369:m1847. https://doi.org/10.1136/bmj.m1847
18. Panda S. Publishing in the time of pandemic: Editorial policy of a dermatology journal during COVID-19. Indian J Dermatol Venereol Leprol. 2020;86(4):337–340. https://doi.org/10.4103/ijdvl.IJDVL_704_20
19. Phillips H. More eyes on COVID-19: Perspectives from History: The need for history in a time of plague. S Afr J Sci. 2020;116(7/8), Art. #8495, 1 page. https://doi.org/10.17159/sajs.2020/8495
20. Callaway E. Will the pandemic permanently alter scientific publishing? Nature. 2020;582(7811):167–168. https://doi.org/10.1038/d41586-020-01520-4.

Peer review – challenges, pitfalls and trust

AUTHOR:
Bridget Farham[1,2]

AFFILIATIONS:
[1]Editor, South African Medical Journal, Cape Town, South Africa
[2]Consultant, WHO African Regional Office, Emergency Preparedness and Response, Health Information and Risk Assessment Programme, Brazaville, Republic of the Congo

CORRESPONDENCE TO:
Bridget Farham

EMAIL:
ugqirha@iafrica.com

HOW TO CITE:
Farham B. Peer review – challenges, pitfalls and trust. S Afr J Sci. 2020;116(9/10), Art. #8796, 2 pages. https://doi.org/10.17159/sajs.2020/8796

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
COVID-19, preprint, medical research

PUBLISHED:
21 September 2020

Just this morning I had reason to think about the whole issue of peer review in our local academic journals. South Africa is in the enviable position on the continent of having a large body of highly regarded and internationally recognised academic institutions. As editor of the *South African Medical Journal* (SAMJ) I have a long tradition of excellence to uphold – sometimes a daunting task. Over the years the SAMJ has evolved and is now a well-respected repository of research and policy papers that are read and cited around the world. During the COVID-19 outbreak in the country, the SAMJ has attracted a large number of submissions from our body of clinicians and researchers, all of whom are happy to submit to a local journal, where I have expedited review and online publication in the interests of keeping people abreast of all the various issues around the current pandemic.

So why was I thinking about peer review specifically? A paper dealing with a particularly important aspect of COVID-19 in our context had been submitted in late June. One of the authors contacted me to enquire about its status, concerned that this potentially important piece of research was taking such a long time to be considered. The reason – waiting for peer review. The SAMJ gives its peer reviewers a limited time in which to review, after which we invite new reviewers. This particular paper had been declined, or simply ignored, by six reviewers. This is not an uncommon situation, unfortunately.

We have approximately 1200 reviewers registered on our website, of whom we use around 200, among whom 150 are regulars and complete reviews within the allocated time or ask for more time if necessary. Wikipedia's definition of peer review is that it is 'the evaluation of work by one or more people with similar competencies as the producers of the work. It functions as a form of self-regulation by qualified members of a profession within the relevant field … and in academia, scholarly peer review is often used to determine an academic paper's suitability for publication'[1].

The pitfalls of peer review are many. One of the biggest is personal bias among others in a field, which can lead to a research submission being rejected simply because the findings do not fall within an accepted paradigm. For those of us who are professional editors rather than academics or, in my case, clinicians, in a particular field, this can be problematic if you are not aware of the various factions in different institutions. It pays to look through the general literature regularly and identify interest groups in particular fields.

There is an increasing move towards pre-publication of papers before they go for formal peer review – the idea being that anyone in the field can look, comment and discuss, potentially adding value to a discussion that would otherwise be limited to perhaps two or three formal reviewers. I have been generally in favour of this practice, having seen many instances where good research has come up against bias that has prevented publication. However, there are pitfalls here as well and the current COVID-19 pandemic has shown these very effectively. In April 2020 a group of Dutch engineers self-published a study that they said showed that the 1.5-m distance rule for people exercising outdoors was flawed.[2] Their premise was that an infected person who exhaled, coughed or sneezed while walking, running or cycling would leave behind them a slipstream of microdroplets. Someone following after such a person would then move through this cloud and potentially inhale the virus. These conclusions were reached by simulating the release of saliva particles from people in motion. They displayed their results as a series of animations and figures in which the cloud of droplets behind a moving person is clearly visible. The simulations led the group to advise that walkers, runners and cyclists should keep a distance of at least 4–5 m behind the leading person while walking in the slipstream, 10 m while running or cycling slowly and at least 20 m when cycling fast. Their argument against waiting for peer review was the 'urgency of the situation and the world-wide crisis'. This led them, in their words, to turn things 'upside down. First research results and communication to the public', then the proposal for funding and only then submission of an article for peer review.

The study went viral. The lead author gave an interview to a Belgian paper and also tweeted about their results. The media picked it up, leading to horrific headlines in, for example, the *Daily Mail*, 'Horrifying simulation reveals the dangers of jogging during the coronavirus pandemic'.[3] Locally, the various forms of 'publication' were spread around Facebook and WhatsApp running and cycling groups (personal observation). A fierce debate started, and other non-academic media picked this up, with 'peer review', much of it from virologists, with some excellent discussions of whether or not the engineers' results could be translated into the real world. One of the best was in wired.com[4], in which Eric Niiler carefully picked the study apart, pointing out that at the time of publication, although we knew that the virus could spread from person to person from coughing or sneezing, we did not know the amount of time that the virus could survive outside the body, or the dose of the virus required to result in infection. At that time there were no published studies of the spread of SARS-CoV-2 from person to person outdoors. A study at the time of 318 outbreaks of three or more COVID-19 patients found that all but one transmission occurred indoors, but that study was also pre-published on medRxIV, a preprint server for health sciences.[5]

More recently, and of far more significance, is the announcement by Russia that they have developed a vaccine against COVID-19. So far no-one has seen a publication, let alone a peer-reviewed publication, on the safety and efficacy of this vaccine. The World Health Organization is currently calling for a 'rigorous safety data review', as the vaccine is currently listed as being in Phase 1 trials, meaning that by most normal criteria, it is nowhere near ready for approval, manufacture and roll out.[6] A search of Pubmed retrieved zero results, showing that not only have studies on the vaccine not been peer reviewed, but they have not even been published. Apparently 20 countries have already pre-ordered more than a billion doses of this vaccine, which by any reasonable criteria is completely unproven.

Here we have, in a nutshell, an example of the importance of peer review in trust. For all the flaws inherent in academic publishing, this is the gold standard that is required before research results can be put into practice.

References

1. Wikipedia. Peer review [webpage on the Internet]. No date [cited 2020 Aug 08]. Available from: https://en.wikipedia.org/wiki/Peer_review
2. Bocken B, Malizia F, Van Druen T, Marchal T. Social distancing v2.0: During walking, running and cycling [document on the Internet]. c2020 [cited 2020 Aug 15). Available from: https://nysoralms.com/wp-content/uploads/2020/04/Social-Distancing-v20_White_Paper.pdf
3. Kekatos M. Horrifying simulation reveals the dangers of jogging during the coronavirus pandemic. Daily Mail. 2020 April 08 [cited 2020 Aug 15]. Available from: https://www.dailymail.co.uk/health/article-8201877/Horrifying-simulation-shows-six-feet-not-jogging-someone.html
4. Niiler E. Are running or cycling actually risks for spreading COVID-19? Wired. 2020 April 14 [cited 2020 Aug 15]. Available from: https://www.wired.com/story/are-running-or-cycling-actually-risks-for-spreading-covid-19/
5. Qian H, Miao T, Liu L, Zheng X, Luo D, Li Y. Indoor transmission of SARS-CoV-2 [preprint]. medRxiv. https://doi.org/10.1101/2020.04.04.20053058
6. WHO wants a safety data review on Russian COVID-19 vaccine. Medical Brief. 2020 August 12 [cited 2020 Aug 15]. Available from: https://www.medicalbrief.co.za/archives/who-wants-a-safety-data-review-on-russian-covid-19-vaccine/

AUTHOR:
Keyan G. Tomaselli[1]

AFFILIATION:
[1]Department of Communication Studies, University of Johannesburg, Johannesburg, South Africa

CORRESPONDENCE TO:
Keyan Tomaselli

EMAIL:
keyant@uj.ac.za

HOW TO CITE:
Tomaselli KG. Peer review is academic citizenship. S Afr J Sci. 2020;116(9/10), Art. #8728, 2 pages. https://doi.org/10.17159/sajs.2020/8728

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
peer reviewing, recognition, para-authors, public benefit

PUBLISHED:
21 September 2020

Peer review is academic citizenship

Academic work has a rapidly increasing half-life while the half-life of academics is becoming a serious problem. As the Editor of *Visual Anthropology,* a very active retiree, recently wrote to the editorial board, on which I serve:

> *Dear Colleagues,*
>
> *… Universities are expecting people to perform or perish … What it means for reviewing is that you Editors might get your best results by pursuing senior or retired experts who are no longer concerned with chalking up points; alternatively, you may have some luck with graduate students, who also may be very up to date on the literatures.*

Globally, academics are wilting under 60–80-hour working weeks, and, not surprisingly, dying prematurely, falling ill or retiring early. Besides, are retirees really still abreast of the current literature?

Some of South Africa's 323 journals are merely aggregators of under-evaluated articles that fill space and enable universities to milk the Department of Higher Education and Training (DHET) incentive. Relentlessly harassed by performance management contracts, even senior academics are submitting half-digested, poorly written, badly referenced and flawed work, and then wondering why competent reviewers respond harshly.

Three categories of reviewer are identified by Ndukuyakhe Ndlovu: (1) the non-responsive; (2) those who never deliver; and of those who do, (3) a failure sometimes to engage with the paper meaningfully.[1] Then there is (4) the report that is calculated to humiliate; and in rare instances, (5) the assessment is designed – no matter the study's potential and relevance – to sink it without trace. Nothing is gained under any of these scenarios.

In contrast, fastidious reviewers and editors are now spending so much time and effort playing the roles of advisor, copy editor, referencing and fact checker that some characterise themselves as para-authors. After all, they are the unsung ones who have rendered publishable initially unpublishable submissions. Yet, the extensive work invested is occasionally seen as obstruction. In one case pertaining to the journal which I edit, a reviewer unselfishly familiarised herself with an obscure topic, generating over 2 days a very helpful two-page report. But the author had just withdrawn the article because of the 'delay'; then requested the report, but did not relay thanks for the time, investment and expertise that the reviewer had expended. For such authors, editors are simply postmasters, and peer reviewers are a time-consuming nuisance.

The epidemic of recent article retractions – especially in the sciences – is indicative of the push to publish prematurely. But few actually perish in the age when evidence, cross-referencing and accuracy are low on the agenda. This is not necessarily fake science but potentially good science managed badly. So bad has it become that a former president of the British Science Association mischievously suggested limiting each academic to but one article annually.[2] This approach would bankrupt South African universities, given their dependence on the well-intentioned DHET publication incentive.

The push to publish is felt by students also. As one told me:

> *Young academics, such as myself, are slowly losing interest in producing high-impact research because the road to an advanced degree is driven by shortcuts and an obvious push by our professors to sacrifice quality and simply produce, produce, produce ... My point is: instead of complaining about how creaky the system is, since we ARE the system, should we not be engaging in constructive dialogues directed at changing the system and making it work for us instead of against us?*

Fortunately, the Academy of Science of South Africa (ASSAf), the National Research Foundation and DHET constitute the national infrastructure of the system. Quality (not quantity), increased research capacity (not depletion through exhaustion) and global competitiveness (not parochial myopia) are the driving criteria of these institutions. These laudable objectives have, however, become lost in the administration of research and institutional need to ensure the bottom line. Peer reviewing is an absent category on performance management templates, even as a form of community engagement unless emphasised by the form-filler.

The exemplar held up by ASSAf's journal evaluation panels is the *South African Journal of Science,* with its full research studies, and its shorter research letters, and front-section commentaries, book reviews and news items. These latter sections are highly read, and sometimes calculatedly controversial, but they are not counted in performance management or DHET annual returns. In other words, there is no 'return' for authors and universities for anything but the discrete incentive-earning article, whether or not it is read, cited or impactful. It is in the commentaries and review essays, however, that the fruitful debates often occur, and which draw the highest citations.

Voluntary editors and peer reviewers are the unpaid cash cows enabling impatient South African university-affiliated authors to feed the subsidy millions into their universities and onward to themselves, depending on internal disbursement policies. That is the institutional upside. The downside is that editors labour after hours, with little or no recognition from many of the institutions that employ them.[3] Are editors just gatekeeping cogs? 'Just write', might be the auditor's instruction; do not edit, do not peer review, do not engage your peers through commentary, book reviews or research letters. Only production-line products – 'accredited' of course – will qualify for research incentives and institutional recognition.

Peer reviewing is not gatekeeping, peremptorily preventing publication or incentive earning. Rather, reviewers are skilled advisors who, while possibly rejecting an article, can be nevertheless helpful for enabling revisions. We

are all learners and should treat each other as such. Moral panics of the kind unleashed from Nicoli Natrass's Commentary[4] may have their place within the non-academic commons, but such ad hominem accusatory responses themselves require rigorous critical discourse analysis – what can be learned from them in terms of ideological positioning? All parties to this debate need to consider that reaching the 'ultimate opinion', to use Peirce's[5] term, involves making sense via a process of semiotic reasoning, and analysing how one came to one's interpretations. [See SAJS Vol 116 Special Issue for debates on Nattrass's Commentary.]

The flip side is: does the author respond appropriately? On occasion, the exact same draft that I rejected for Journal A, notwithstanding the helpful comments offered, is submitted to Journal B, and again finds its way to me to review, and again with regard to Journal C, which identifies me as a reviewer also. Eventually, this unrevised article will be published in Journal Non-entity, also 'accredited'. And, hiding in plain sight, is the very high continuing incidence of plagiarism within the 17 South African management journals.[6] That no outcry has occurred in this recurring instance is an indicator of a catastrophic failure of peer review and disciplinary and institutional accountability, with attendant costs to the Treasury.

'How-to-get-published workshops' should be complemented with training in peer review. The 'yin' cannot work without the 'yang'. Also, in the age of big copyrighted data, reviewers should additionally be able to scrutinise the data bases on which studies are predicated and be privy to source codes of customised software.[7]

The best reviewers are those who (1) critically read and constructively engage the submission; (2) offer helpful comments to enable the author to improve the study; (3) refer authors to cognate studies that would strengthen (or contest) their own arguments and findings in a holistic mapping of the topic; and (4) submit reports on time, especially given the half-life of knowledge. Finally, (5) peer review assists in the much vaunted objective of 'de-colonisation'.[8] As Ndlovu[1] observes, when African-based scholars evade peer reviewing duties, editors have to rely on their overseas colleagues who may be insufficiently familiar with local contexts.

Some South African authors are mystified and resentful when engaging in extended dialogue with editors, reviewers and copy editors, sometimes over many drafts, over many months. Such collaboration is part of the process, and on occasion I read sentences that I wrote in a report now being used verbatim by an author without attribution to the anonymous reviewer! Thus, do peer reviewers or para-authors voluntarily cede their intellectual property to someone else who cannot acknowledge reviewer generosity where a blunt blind procedure is applied.

Another bugbear is when an article is published exactly as first submitted notwithstanding extensive reviewer reports and recommendations. The question arises: did the editor pass the report onto the author or not? If not, why not, and why expose the now published author in a vulnerable situation? Or, totally contradictory reports are passed on to authors with a mere request that the criticisms simply be addressed.

Preprints became the order of the day during the COVID-19 pandemic. However, in the rush to out-Trump Trump's Fox TV-led fake science, pseudo-pharmacology and rhetorical statistical denialism, impatient scientists risked adding to the retraction factor. South Africa, especially, has a wretched history of promoting sham reasoning and dodging peer review as occurred during the era of HIV/Aids denialism. When evaluation protocols are evaded, uncontrolled medical experimentation results, making nonsense of ethical criteria and public health, not to mention scientific validity. The dilemma: fast-breaking information is needed when health emergencies arise. In such cases, peer review generates 'a rapid-results process', and most COVID-19 preprint websites did institute 'appropriate systems' (with thanks to public health communication specialist Warren Parker for this observation). In contrast, thanks to opportunistic COVID-19 Trumpian politicisation, the Chinese introduced party political review, in addition to peer review.

Peer review is not perfect, but it offers the best current practice. When we fail our peer-reviewing duties, we fail ourselves, our disciplines, our institutions and the public. When university managements fail to acknowledge the fundamental value of peer review, they imperil science, and conceptual and methodological progress.

For ASSAf, an article attains value when it factors *readers* and social usefulness of research into the national equation. Its National Scholarly Editors' Forum (NSEF) regularly debates different models of peer review and places social value – rather than just the product and metrics – as a key objective of academic citizenship. Debate on academic research practice is encouraged at NSEF meetings that bring editors together from the public, private and university sectors. Journal editing, peer review and academic citizenship ensures a holistic and community-oriented approach to our work where agreed rules of engagement are followed. The Nattrass affair has reinstalled commentaries back on the debating agenda, DHET disincentives notwithstanding. Peer review is a fundamental communal practice of critical academic citizenship, and it contributes towards improving science for public benefit. It should not be an unrecognised add-on done in one's spare time. But for academic auditors, peer review is not itself an income-earning activity. There's the rub.

We all need to do peer review, properly. Our half-life will otherwise be fruitless. Getting published is just one component of academic citizenship – one that cannot function without review procedures. But the practice should never be reduced to box ticking. For retirees, reviewing keeps them intellectually active. For science, peer review is the fuel that drives the system. For authors, peer review is quality control, and for readers, the practice is an assurance of reasonable validity. For the public, peer review, especially in the medical sciences, could be the difference between life and death.

References

1. Ndlovu N. Wheels coming off: Critically assessing the peer review process in southern African archaeology. S Afr Archaeol Bull. 2018;73(208):79–81.
2. Grove J. Restrict researchers to one paper a year, says UCL professor: Uta Frith says limit would force academics to focus on quality over quantity. Times Higher Education. 2019 November 28. Available from: https://www.timeshighereducation.com/news/restrict-researchers-one-paper-ayear-says-ucl-professor#node-comments
3. Tomaselli KG. Perverse incentives and the political economy of South African academic journal publishing. S Afr J Sci. 2018;114(11/12), Art. #4341, 6 pages. https://doi.org/10.17159/sajs.2018/4341
4. Nattrass N. Why are black South African students less likely to consider studying biological sciences? S Afr J Sci. 2020;116(5/6), Art. #7864, 2 pages. https://doi.org/10.17159/sajs.2020/7864
5. Peirce CS. Collected papers of Charles Sanders Peirce. Volume 7. Cambridge, MA: Belknapp Press of Harvard University Press; 1966.
6. Thomas A. Plagiarism in South African management journals: A follow-up study. S Afr J Sci. 2019;115(5/6), Art. #5723, 8 pages. https://doi.org/10.17159/sajs.2019/5723
7. Heathers J. The Lancet has made one of the biggest retractions in modern history. How could this happen? The Guardian. 2020 June 05. Available from: https://www.theguardian.com/commentisfree/2020/jun/05/lancet-had-to-do-one-of-the-biggest-retractions-in-modern-history-how-could-this-happen?CMP=Share_iOSApp_Other
8. Tomaselli KG. Indeterminacy, indigeneity, peer review and the mind–body problem. Junctures: The Journal for Thematic Dialogue. 2019;20:87–102. http://dx.doi.org/10.34074/junc.20087

AUTHOR:
Johann Tempelhoff[1]

AFFILIATION:
[1]Research Group for the Cultural Dynamics of Water in Southern Africa, School of Basic Sciences, Vaal Triangle Faculty, North-West University, Vanderbijlpark, South Africa

CORRESPONDENCE TO:
Johann Tempelhoff

EMAIL:
johann.tempelhoff@nwu.ac.za

HOW TO CITE:
Tempelhoff J. Peer review in academic journals: A Pandora's box. S Afr J Sci. 2020;116(9/10), Art. #8783, 2 pages. https://doi.org/10.17159/sajs.2020/8783

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
academic publishing, gatekeeping, cooperation, mentoring, recognition

PUBLISHED:
21 September 2020

Peer review in academic journals: A Pandora's box

A recent open webinar – a regular public digital platform for academics to communicate in the time of COVID – hosted by the Academy of Science of South Africa (ASSAf) on 31 July dealt with the peer review process of academic journals.

Although it remains something of a Pandora's box at times, peer review is undoubtedly a key component in the rituals of academic publishing. Although often fraught with surprises, conflict, and lively debate, it is peer review that ensures sound quality.[1]

Currently, the South African academic publishing industry is in a growth phase. The number of manuscripts submitted is increasing, with many written by a new generation of academics eager to make their mark in their chosen fields of expertise. But alas, many authors encounter obstacles in getting their material beyond the peer-review phase.

For obvious reasons there is considerable interest among academics to find out more about the peer-review process – evident by the over 300 people who registered in advance to attend the digital webinar.

Presenters included Prof. Lucienne Abrahams of the University of the Witwatersrand and editor of *The African Journal of Information and Communication*; Dr Salmina Mokgehle, a researcher at the Agricultural Research Council and Associate Editor Mentee at the *South African Journal of Science* (SAJS); Prof. MP Sebola, a widely published local and international academic in the field of management sciences at the University of Limpopo; and Dr Andy Carolin, a senior lecturer at the University of Johannesburg and a regular peer reviewer in the fields of English literature, gender and cultural studies.

Abrahams, who facilitated the proceedings, stressed in her opening statement that creativity is important. An innovative piece of writing that is informative and beyond the run-of-the-mill type of material often gets published. Her advice to prospective authors was to gain insight into research methodology and writing skills in areas of investigation beyond their own fields of specialisation. She made the point that by collaborating in research groups, authors are exposed to valuable, innovative ideas which they can then apply in their own research and reporting.

Panellists were well aware of the typical problem authors have with 'exclusivist' and 'closed shop' journal peer reviews and speakers acknowledged that specific approaches and/or trends are pervasive in some disciplines and that this could lead to journals losing contributions from a new generation of academic writers.

Carolin acknowledged that the role of the peer reviewer is pivotal because many reviews operate in a contested field. There would be diverse views and reviewers might well lack consensus on an article. Editors would then be obliged to find an additional reviewer. However, more often than not, the journal's editor would have to step in to make the ultimate decision on the material in question.

Reviewers, Carolin told the participants, are, after all, ordinary people. They are fallible. Their only advantage over the next reviewer is their specific field of knowledge. If a reviewer shows the correct attitude, a sensible assessment of a manuscript can be reached. Carolin's classification of peer reviewers included the 'monster' (the one who obstructs output); the gatekeeper (guarding the disciplinary domain in a dogmatic mode of thinking); and the 'desirable peer reviewer' who fits into the role of becoming an indirect contributor by making constructive suggestions on a manuscript. It is this latter attitude that contributes to a process of meaningful knowledge production.

There is no place for bad scholarship in the publishing realm. At the same time, there must be a sense of generosity towards the author, as well as to the peer reviewer. According to Carolin, it is sound, sensible editorial management that secures appropriate reviewers.

Sebola, an experienced and much-published academic author, explained that it remains difficult to understand why a local journal would decline his work, while the same material sails through peer reviews of highly respected international journals. His advice was for authors to carefully identify the appropriate type of journal for the material an author may have on offer. South Africa is currently moving into a phase of significant postgraduate research output, according to Sebola, and this new generation of academics is eager to make its mark.

On the editorial side, the SAJS has introduced an Associate Editor Mentorship Programme. Mokgehle explained that appropriate mentoring has given her the opportunity to communicate with authors of highly promising manuscripts and allowed her to engage in a systematic phase of coaching in writing and editing.

From the discussion, it was evident that the SAJS strategy has merit. When promising young academics are brought into the editorial management system, they come from new networks and have fresh insight into emerging fields of knowledge. Given the appropriate editorial mentoring, the mentee can make a substantial contribution to the further absorption of, for example, indigenous African knowledge systems in creative academic writing. Increasingly, younger editorial members are becoming familiar with current local and international trends in numerous fields of research.

From Mokgehle's discussion, it was evident that SAJS's young editorial members soon became familiar with the phenomenon of a range of views shared by reviewers. The best approach, it seems, is to secure a review from the right expert in the specific field and to make sure that the material is compliant with the guidelines of the journal.

As mentioned, South Africa is currently on the cusp of a higher production rate of completed postgraduate studies that should naturally correspond to an increasing rate of published articles. However, participants in the webinar

discussion pointed out that the postgraduate output did not always lead to reporting in academic journals. Both the panellists and participants argued that academics have busy schedules. Apart from teaching they also have their own research and pressing administrative tasks, along with the customary load of postgraduate students. They do not always have the time to provide additional guidance on the manuscripts that arrive on their desks to review.

On the editorial side, it appears that the ever-present problem is the inability to have enough, or the right type of, peer reviewers. The speakers as well as the webinar participants favoured suggestions that peer reviewing academics should get recognition from their managers for the time and effort spent on peer reviews and for serving on journal editorial boards. Peer reviewing should feature in their performance management assessments.

In fact, Abrahams was of the view that peer-reviewing responsibilities should be taken into consideration when promotions are made. Furthermore, ratings institutions such as the National Research Foundation of South Africa should take note of editorial services rendered by academics.

In some quarters there were calls for peer reviewers being paid for doing reviews. However, during the open discussion, there was consensus that the peer review should be seen as a free service rendered by individual academics who could be relied on to make sound judgements as a community service to other experts.

There were questions about young academics submitting material with, or without, the support of their postgraduate study leaders. Some speakers hinted that study leaders were not always supportive of article writing. One editor insisted that study leaders should play a more supportive role in the guidance of their students, especially in the field of academic publishing. With study leaders typically taking charge of mentoring the student from the writing phase to the final manuscript submission, the success rate of the manuscript could increase significantly.

The elephant in the room remained a sense of cynicism about the fact that material submitted seldom reached the phase beyond a first peer-review round.

According to Sebola, the Association of Commonwealth Universities and the management of South Africa's universities are well aware of the value of next generation scholarship. He stressed the point that editors of journals need to facilitate the process of absorption and growth. Personally, he favoured the arrangement that the approval of an article would not be the exclusive, privileged domain of the editor. It had to be a collective editorial decision.

Moreover, there had to be more opportunities for authors to learn from the peer-review process. Reviewers, Sebola pointed out, need to provide detailed expositions of what needed to be done to improve the quality of a manuscript.

Participants in the open discussion agreed that editors and reviewers working with large numbers of manuscripts, seldom had the time to give detailed and ongoing support on the editorial content of articles. It was felt that the responsibility had to be shared by the author and postgraduate study leader. The study leader was perhaps in the best position to do valuable coaching.

Some good ideas were discussed. There were suggestions of using Publons (a commercial website that provides a free service for academics to track, verify, and showcase their peer review and editorial contributions for academic journals).[2] Also thought provoking was the question: Who does the young academic go to for support when doing a first peer review for a journal? Much can be said for a future author as reviewer, equipped with some basic postgraduate skills on reviewing practices.

Owing to time constraints, discussion leaders were unable to make disclosures about the complex process of digital publication in today's virtual academic journal editorial office. In recent years, thanks to digital technologies, the publishing process has sped up. Gone are the days when manuscripts, typed or handwritten, and sent in hard copy and by registered mail, would be subject to typesetting and proofreading before being published.

Nowadays, publishing is fast, complex and multi-faceted. Many processes happen simultaneously. Manuscripts are subjected to digital similarity checks to prevent plagiarism. Algorithms monitor the impact of published texts. The quantitative measure has replaced many uncertainties of the qualitative processes of former times.

However, peer review, it seems, remains of crucial importance in successful publications. It is the task of the editor, or members of an editorial team, to draw a distinction between the good, the bad, and the 'rude' peer review. The editorial team has to consider the reviews from the perspective of an academic commitment that speaks to integrity and ethics.[3] Reviewers who are disrespectful to authors, be they young academics or old-fashioned veterans, should not feature prominently on the name lists of good journals.

Any experienced journal editor would concede that a careful manuscript is more likely to go all the way to publication than a shoddily written text. Manuscripts that make the grade are usually those that have been read first by an experienced and trusted colleague, before being passed on to a skilled language editor and submitted with the greatest respect for the editorial guidelines of the identified journal. After all, there is pride at stake in a good piece of writing. Few authors favour publishing their work anonymously.[4] A neat article speaks of self-pride.

Even academic journals themselves are subject to internal and external oversight and peer review. Reviews by ratings entities determine their status. Editorial board members tend to be large in number and they work in close collaboration with technology systems administrators. Nowadays globalisation also plays an important role. The peer review of a good, local-content article often reaches the desktop of an expert reviewer in a foreign country.

Editors, confident of their submitted material, take pride in sharing local work of outstanding quality with colleagues elsewhere. In the field of natural science, peer review is said to be the critical objective way of ensuring that there is transparency and trust in the findings that are shared by a number of specialists in a particular field.[5] This is also the view upheld in the social sciences and humanities.

Peer review, although somewhat of a Pandora's box for many authors and editors, is, however, bound to remain a much appreciated and respected element of academic publishing.

References

1. Wingfield B. The peer review system has flaws. But it's still a barrier to bad science. The Conversation. 2017 September 20. Available from: https://theconversation.com/the-peer-review-system-has-flaws-but-its-still-a-barrier-to-bad-science-84223
2. Smith D. Will Publons popularize the scientific peer-review process? BioScience. 2016;66:265–266. https://doi.org/10.1093/biosci/biw010
3. Mavrogenis A, Quaile A, Scarlat M. The good, the bad and the rude peer-review. Int Orthop. 2020;44:413–415. https://doi.org/10.1007/s00264-020-04504-1
4. Anon. Would scientists be willing to write anonymous papers? Curr Sci. 2016;111:250–251.
5. Baumberg J. The secret life of science. Princeton, NJ: Princeton University Press; 2018. p. 55–90.

AUTHOR:
Robin Crewe[1]

AFFILIATIONS:
[1]Senior Research Fellow: Centre for the Advancement of Scholarship, University of Pretoria, Pretoria, South Africa
[2]Chair: Committee on Scholarly Publishing in South Africa, Academy of Science of South Africa (ASSAf), Pretoria, South Africa

CORRESPONDENCE TO:
Robin Crewe

EMAIL:
robin.crewe@up.ac.za

HOW TO CITE:
Crewe R. Assuring the quality of scholarly South African journals: An experiment in journal peer review. S Afr J Sci. 2020;116(9/10), Art. #8795, 3 pages. https://doi.org/10.17159/sajs.2020/8795

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
grouped peer review, funding formulae, journal databases

PUBLISHED:
21 September 2020

Assuring the quality of scholarly South African journals: An experiment in journal peer review

> **Recommendation No 5:** that ASSAf be mandated jointly by the Departments of Education and Science and Technology to carry out external peer review and associated quality audit of all South African research journals in 5-year cycles, probably best done in relation to groups of titles sharing a particular broad disciplinary focus, in order to make recommendations for improved functioning of each journal in the national and international system.
>
> *Academy of Science of South Africa (ASSAf)*[1]

The introduction of the research output incentive scheme for public higher education institutions (HEIs) in the 1980s[2] was premised on two principles – the first was that rewarding outputs would direct resources to those institutions that were producing graduates at the master's and doctoral levels, and whose staff and students produced published scholarly work in recognised journals, conference proceedings and books. The incentive scheme was seen as a mechanism to encourage improvement in both the productivity and the quality of the research produced by the public HEIs in South Africa. With the introduction of the New Funding Framework for HEIs in the 2004/2005 financial year, the way in which institutions received funding was explicitly laid out[3] and research was funded as a separate category with its components funded through awarding units with a particular monetary value.

The component of the formula for research funding that will be explored here relates to the recognition of scholarly publications in journals only. Scholarly publications in books and conference proceedings were assessed by a committee established by the Department of National Education (at that time) that decided on the subsidy units to be allocated to these categories of submission. Journal outputs were handled differently, with the Department indicating that only journals listed in the Institute for Scientific Information (ISI) databases and the International Bibliography of Social Sciences (IBSS) would be automatically recognised. South African journals that were not listed by the ISI or IBSS had to apply for inclusion in the Department of National Education's and then subsequently the Department of Higher Education and Training (DHET)'s list of accredited South African journals (LASAJ). The quality of journal outputs was assessed using the ISI and IBSS databases as a proxy for acceptable quality while South African journal titles that were not included in these lists had to apply to the Department for inclusion in the LASAJ.

Just prior to the change in the formula for the funding of universities, the Department of Science and Technology (DST) decided in 2001 that it would close the Bureau for Scientific Publications which had previously provided funding for a group of South African research journals, and it contracted the Academy of Science of South Africa (ASSAf) to review scholarly publishing in South Africa and to make recommendations to DST and DHET regarding this matter. The ASSAf report[1] that was delivered in 2006, and its fifth recommendation (shown above), proposed the idea of subjecting journals to peer review.

The key element of this recommendation was the systemic peer review and periodic audit of all South African journals and particularly those included in the LASAJ. At the time that this recommendation was made it was entirely novel in that it proposed the peer review *of* journals in a way that was analogous to the peer review undertaken *by* journals of the material submitted to them. It was the first time that a peer-review process was being proposed to assess the quality of material in, and editorial processes of, journals. The underlying thrust of the approach was to encourage accountability for the quality of content and also make developmental recommendations directed at improving the quality of the journals reviewed.

Following the acceptance of the ASSAf report, ASSAf undertook the task of implementing the recommendations of the report and the ASSAf Council established the Committee on Scholarly Publishing in South Africa (CSPiSA) which had oversight of ASSAf's Scholarly Publishing Programme (SPP). It was the staff of the SPP that were given the task of making journal peer review a reality.

Introduction of peer review of journals

An explicit call for the peer review of South African scholarly journals was made in Recommendation 5 of the ASSAf report[1] with the idea that this be mandated by the Department of Education and DST which were the relevant government departments at that time. Both departments endorsed the idea and indicated that the process could be used to manage South African journals on the LASAJ. This proposal was then discussed with the newly established National Scholarly Editors' Forum (NSEF) and the editors present supported the introduction of this process and were consulted regarding both the process and documentation that would be used. The outcome of this consultation was that guidelines for the peer review of journals were developed and introduced.[4]

There is a clear distinction between the process that ASSAf uses to review journals and makes recommendations regarding their inclusion in the LASAJ and/or their suitability for the SciELO South Africa platform[5], from that used by other database services such as Web of Science (successor to ISI), IBSS, Scopus and others.

The currently DHET recognised databases outside of South Africa are the Norwegian List, Web of Science, IBSS, and Scopus. These four databases try to ensure that there is some form of quality assurance in relation to the items that are included. However, the form of quality assurance employed usually consists of a set of technical criteria that the journal needs to address in order to be listed. Individual journals may be delisted if they no longer adhere to the stipulated criteria. In addition, some of the companies maintaining these databases have initiated the use of bibliometric data to screen journals for acceptability. An example of the criteria that are used by Elsevier for the Scopus database are:

> *To be considered for review, all journal titles should meet all of these minimum criteria:*
>
> - *Consist of peer-reviewed content and have a publicly available description of the peer-review process;*
> - *Be published on a regular basis and have an International Standard Serial Number (ISSN) as registered with the ISSN International Centre;*
> - *Have content that is relevant for and readable by an international audience, meaning: have references in Roman script and have English language abstracts and titles;*
> - *Have a publicly available publication ethics and publication malpractice statement.*

In addition to this list of minimum requirements, there is also a set of metrics that the journals have to meet in order to be included or remain on the list.[6] In addition, Scopus has 'The Scopus Content Selection and Advisory Board (CSAB) [that] is an international group of scientists, researchers and librarians who represent the major scientific disciplines.' The way in which this CSAB operates and how it is constituted is not clearly spelt out.

The peer review of journals undertaken by ASSAf is different in nature, rigour and transparency from the procedures that appear to be used by the other journal databases. The key element of this difference is the process of peer review and the way that it is implemented which follows a standard methodology that has been refined by ASSAf over a number of years and provides the outcomes of its reviews in the form of publicly available reports.

ASSAf journal peer review process

The review of a group of journals is initiated by the CSPiSA, which then seeks a group of suitably qualified individuals to constitute a peer-review panel to undertake the review of the chosen set of journals. The CSPiSA solicits the nomination of individuals who are recognised scholars in the disciplinary fields of the journals to be reviewed. In addition, individuals with expertise in journal publishing as well as one individual who is from an unrelated field of study are included in the selection. Generally, 10–12 individuals are nominated and the panel is then constituted of 6–8 members who agree to participate, and whose membership is approved by the ASSAf Council.

The panel elects its own chair and manages, with support from staff of the SPP, the production of a report on the journals being reviewed. The process entails the identification and appointment of the peer reviewers who will review the issues of the journals that have been selected for review. Two or three peer reviewers are appointed for each journal. In addition to the peer reviewers' reports, information is obtained from the editor/editorial board of the journal that details the way in which the journal implements the 'Code of Best Practice in Scholarly Journal Publishing, Editing and Peer Review'[7] and other details of the journal's practices and business model. In addition, bibliometric data on the performance of the journal may be obtained.

The peer reviewers' comments and recommendations regarding a journal, together with the other data obtained about the journal, are considered by the review panel and a report on the journal is formulated. The report on each journal is based on the data collected about the journal together with the peer reviewers' comments. This synthesis is prepared by the panel and represents their consensus view of the report on the journal.

Once the report on a particular journal has been approved by the review panel, it is sent to the editor/editorial board of the journal for comment. The request for comment is not regarded as an opportunity for rebuttal or explanation of the findings of the review panel, but rather an opportunity for the journal's editor to point out any errors of fact in the report or to draw the attention of the panel to any oversights that may have occurred. The panel considers the comments that it receives to this draft of the report and may make appropriate revisions if these are considered to be necessary.

The panel is also asked to make recommendations about the following:

1. Should the editor/publisher of the journal be invited to consider placing the journal on the SciELO South Africa platform if it meets the criteria for inclusion?
2. Make a recommendation to DHET regarding the suitability of the journal for LASAJ.
3. If the recommendation in (1) and/or (2) is negative, then guidance should be provided about what the journal needs to do to meet the criteria for inclusion.
4. Any additional recommendations for improvement and enhanced functionality.

Reports on all the journals considered in a particular disciplinary group are then consolidated into a single report which is prepared for publication. Prior to publication, the report is reviewed by the members of CSPiSA and if they are satisfied with its contents, the report is referred to the ASSAf Council for approval. Once approved by Council, the report is published and publicly disseminated, and copies made available for downloading from the ASSAf website.

The data in Table 1 indicate that over a period of a decade, the various review panels have managed to review and make recommendations about journals covering a broad sweep of disciplines. An explicit set of quality assurance mechanisms for assessing the quality of South African journals has been provided. Indeed, in a recent report on scholarly publishing in South Africa[8], the following observation was made: 'We need ongoing analyses of SA publication practices to identify cases of questionable publication and again to alert the DHET and university research offices to such practices.' The report goes on to comment: 'Amongst a large number of interesting and relevant findings were disturbing indications of predatory publishing and questionable editorial practices.' The peer review of journals undertaken by ASSAf provides a level of oversight for South African journals that can mitigate the effects of some of the dubious publication practices identified. The fact that these practices are not more widespread amongst South African journals may at least partly be attributable to the advice provided by the National Scholarly Editors' Forum and the role that journal peer review has played in monitoring journal performance and encouraging reform where deficiencies are identified.

Table 1: The group peer review reports on scholarly journals in particular disciplines that have been carried out to date. The year of publication of the reports is listed. All are openly accessible from the ASSAf website.

Title of report	Year
Report on Peer Review of Scholarly Journals in the Agricultural and Related Basic Life Sciences [link]	2010
Report on Peer Review of Scholarly Journals in the Social Sciences and Related Fields [link]	2010
Report on Grouped Peer Review of Scholarly Journals in Religion, Theology and Related Fields [link]	2013
Report on Grouped Peer Review of Scholarly Journals in Health Sciences and Related Medical Fields [link]	2014
Report on Grouped Peer Review of Scholarly Journals in Law and Related Legal Fields [link]	2014
Report on Grouped Peer Review of Humanities Part 1: Literature Group Classics, Literature and Languages [link]	2015
Report on Grouped Peer Review of Scholarly Journals in Architecture, Built Environment and Engineering [link]	2018
Report on Grouped Peer Review of Scholarly Journals in Humanities II – Visual and Performing Arts [link]	2018
Report on Grouped Peer Review of Scholarly Journals in Communication and Information Sciences [link]	2019
Report on Grouped Peer Review of Scholarly Journals in Education [link]	2020

References

1. Academy of Science of South Africa (ASSAf). Report on a strategic approach to research publishing in South Africa. Pretoria: ASSAf; 2006. http://dx.doi.org/10.17159/assaf/0038
2. Pouris A. Effects of funding policies on research publications in South Africa. S Afr J Sci. 1991;87:78–81.
3. Steyn AGW, De Villiers AP. Public funding of higher education in South Africa by means of formulae. In: Review of higher education in South Africa. Pretoria: Council on Higher Education; 2007. p. 11–51. Available from: https://pdf4pro.com/view/review-of-higher-education-in-south-africa-28a64b.html
4. Academy of Science of South Africa (ASSAf). Periodic peer review of South African scholarly journals: Approved process guidelines and criteria [webpage on the Internet]. No date [cited 2020 Aug 17]. Available from: https://www.assaf.org.za/index.php/programmes/scholarly-publishing-programme/peer-review-panels
5. ASSAf. SciELO South Africa [webpage on the Internet]. No date [cited 2020 Aug 17]. Available from: https://assaf.org.za/index.php/programmes/scholarly-publishing-programme/open-access-scielo-south-africa
6. Elsevier. Content policy and selection [webpage on the Internet]. No date [cited 2020 Aug 17]. Available from: https://www.elsevier.com/solutions/scopus/how-scopus-works/content/content-policy-and-selection
7. Code of Best Practice in Scholarly Journal Publishing, Editing and Peer Review [webpage on the Internet]. c2018 [cited 2020 Aug 17]. Available from: https://sites.google.com/view/assaf-nsef-best-practice
8. Academy of Science of South Africa (ASSAf). Twelve years later: Second ASSAf report on research publishing in and from South Africa. Pretoria: ASSAf; 2019. https://doi.org/10.17159/assaf.2018/0030

AUTHOR:
Phila Msimang[1]

AFFILIATION:
[1]Department of Philosophy, Stellenbosch University, Stellenbosch, South Africa

CORRESPONDENCE TO:
Phila Msimang

EMAIL:
MsimangP@sun.ac.za

HOW TO CITE:
Msimang P. Lessons in our faults: Fault lines on race and research ethics. S Afr J Sci. 2020;116(9/10), Art. #8449, 3 pages. https://doi.org/10.17159/sajs.2020/8449

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
race, research ethics, racial classification

PUBLISHED:
29 September 2020

Lessons in our faults: Fault lines on race and research ethics

Addressing the question of race in scientific research

Since the publication of the now infamous paper 'Age- and education-related effects on cognitive functioning in Colored South African women' by Nieuwoudt and colleagues where they claim the cognitive functioning of coloured women is defective in some ways,[1] there has been renewed doubts about the legitimacy of race in research at Stellenbosch University and the controversy has become a matter of concern for universities and research ethics boards across the country. Because of the harms the Nieuwoudt paper has caused, the question of the use of race in research has re-emerged as a central concern. What researchers want to know is how to assess whether race is relevant to some question(s), how to use racial classifications if race is relevant, and when to leave race out.

The attention which this paper has received is perhaps related to how it is a caricature of a more common and mundane problem about the status of race in research. Considered responses to the paper have centred on the question of how we are to use race in academic settings if at all. The Nieuwoudt paper is a caricature of the problem because the study it presented was fundamentally misguided, it relied on racist stereotypes in a quite explicit fashion, its claims about the homogeneity of the group it studied are patently false (to bolster their claim, they cited a paper that actually makes the opposite claim), and it did not meet basic scientific standards in both its claims and their assessment.[2] Because of its basic scientific failings, some academics have asked how the Nieuwoudt paper managed to pass peer review and how the research protocol managed to attain ethics approval in the first place.[3] A speculative conjecture in this regard is that the reviewers themselves may have held uncritical beliefs about race that made them overlook what, for many researchers and concerned South Africans, were glaringly dubious and problematic claims about race.[3] After a year-long investigation at Stellenbosch University, it was finally announced that 'the article was not aligned with the Research Ethics Committee (REC) approved protocol', meaning that it is arguable that this research did not follow the recommendations of the REC.[4] That the article was soon retracted by the journal for its lack of scientific merit suggests that there was a failure of quality control at the level of the peer review that the journal recognised on a second assessment.[5]

The more general and mundane problem this points to, both locally and internationally (the paper was published in an international journal of good standing), is that this publication may be indicative of a lack of understanding about how to manage the variable of race and its possible uses. It is for this reason that the Nieuwoudt paper inspired numerous symposia and workshops about race and research on the various campuses of Stellenbosch University. Notable are the events held by the University Senate, the Department of Psychology, and the Medical Faculty. Out of the event held by the Medical Faculty came an edited collection, *Fault Lines*.[6] This edited collection aims to tackle the problem of race in research head-on.

The book is a welcome beginning to treatments of race in research and opens the door to more critical discussions about the issues it explores. A glaring gap and fundamental weakness of the collection is the absence of dealing with the material and economic aspects of race and racism despite its historical lens. The book has very little to say about the point of white supremacist ideology through colonialism and apartheid being used to justify dispossession and to ensure the material racial domination and material inequality from which the country still suffers. This may make it seem as if the problem of the continued production of racist research is disconnected from the problem of the continued reproduction of a systemically racist society in not only the academy but in the economic domain as well. Addressing the issues of the material bases of race and racism directly would have gone a long way to further contextualising and illuminating the conditions which produce or afford these research products and the broader social power relations to which they belong and from which they manifest.

The resultant scope and strength of this collection is how it gives some historical context to the use and abuse of race in different research settings, especially in how research ethics have been contravened in respect to race around the world. Importantly, this book provides specific guidelines on the use of race in research in the South African setting. Given that the aim of the book is to help us understand race in research, the collection feels incomplete as it leaves out a central aspect to understanding race – the material and economic – and how that then plays out in our research activities. Nevertheless, this collection is a useful introduction to the debate that can be supplemented with works about the material bases of and incentives for the continuation of racism in South Africa.

Not all contributions in the book endorse racial classification or its use. Some contributors outline an orientation that challenges 'race thinking' and its place in academics, politics, and in social and personal identity. It is unclear how the tensions between the problematic nature of race and race's employment in research is to be resolved under those considerations, but they do open up a space to explore many interesting questions.

Other contributions in the collection argue for critical uses of race, which often accompany an understanding of race as real and having a justifiable place in research. One of the editors of the collection, Johnathan Jansen, puts the case like this:

> *Not all studies of race are socially regressive. Put differently, there are critical studies of race that should not be denied ethical clearance simply because the study is about race. The point of departure for such critical studies of race is not that race is real in the sense of it being a biological or cultural essence; rather, the starting point for such studies is that race is a social category whose functions need to be unmasked.*[7]

I will further comment on three chapters in the collection that I believe hold some important lessons for the use of race as a variable in research – both in terms of thinking about what race is in general and in thinking about if race is relevant to specific scientific studies.

From genetics to race and ethnicity in the human sciences

After an introductory overview, the first chapter of the book takes us through the basics of human genetics and human biological diversity. This chapter, 'The Role of Genetics in Racial Categorisation of Humans'[8], addresses the question of whether race is biologically real. This chapter is penned by two geneticists, Soraya Bardien-Kruger and Amica Müller-Nedebock, from Stellenbosch University.

They make the case that race is not biologically real, and recount the standard reasons related to the apportionment of genetic diversity across our species and how genetically similar people are to one another. Much attention in this contribution is paid to sketching the basics of human genetic diversity and the current consensus about individual- and population-level differences between humans. This makes this chapter especially useful as an introduction to the relationship between race and biology. The authors make their case by explaining why race is not a valid taxonomic category, why race does not represent human genetic diversity well, and why race can sometimes correlate with some clinically or medically relevant information despite race not being biological. The authors do this in a lucid non-technical style that makes it easy to follow even if you do not have a background in genetics.

One reservation I had with this chapter is its treatment of ethnic groups as biological categories. Although ethnic groups may be more reliable in a number of respects in clinical settings and generally have a closer relationship with genetic ancestry than races do, ethnic groups are still social groups and are not biological categories.[9,10] The use of ethnic groups is one way approximations to biogeographical or genetic ancestry are made, as the authors describe, but ethnic groups are not themselves units of ancestry – ethnic groups are not ancestry groups at any level of biological population. This means that ethnicity is separate from genetic ancestry.

All that one needs to do to understand why ethnicity and genetic ancestry are separate is to think about how ethnic classifications operate. Consider the case of 'trans-ethnic' adoption between, say, English parents and an Afrikaner or French child, or a Tsonga family and a Zulu or Shona child. The child's ethnic group can change, yet their genetics remains the same. Even as an adult, the ethnic group someone is said to belong to is heavily dependent on the aspects of their heritage or acquired culture with which they are associated, and this heritage or culture is not strictly tied to specific relations of a biological nature. Ethnic classifications, like racial classifications, are social categories, despite any correlations they may have in some clinical and medical research settings.[11] Racial and ethnic categories, along with all of their surrogates, should not be used uncritically as uncritical uses of these categories can be misleading and have dire consequences that can be quite direct for patients in clinical contexts.

This difference between race, ethnicity and genetic ancestry is made clear by Jimmy Volmink, Lynn Hendricks, Lindokuhle Mazibuko and Leslie Swartz in their contribution 'Race and Health: Dilemmas of the South African health researcher'.[12] They distinguish genetic ancestry from social classifications like race and ethnicity and explain how they operate differently. They explain:

> *Unlike race or ethnicity, which is concerned with how a person fits into a particular group, genetic ancestry focuses on how an individual's history has unfolded – essentially, how his or her ancestors moved and mated. Someone's self-identified or assigned race or ethnicity may therefore differ considerably from data computed using AIMs [*ancestry informative markers*], and may also reflect multiple ancestral origins.*[12]

The concern in this chapter is with the understanding and use of race in health and clinical research settings. The authors make a case for a nuanced approach to the use of racial terminology that takes into account the various environmental and genetic influences that could be at play in any particular health setting. In the same breath, they warn of the dangers of 'clinical racial profiling' and the dangers inherent in profiling in health-care settings more generally. Because the use of race and ethnicity as markers of biological variation is misleading, they discourage their use as guides to such variation, especially as it could undermine patients' quality of care and afford the neglect of investigating or identifying structural pathways to certain health outcomes. It is in light of these considerations that the chapter looks at the use of racial, ethnic, and ancestry related terminology in a year's worth of publications from the medical faculty at Stellenbosch University. In this, they debate the need to keep to international standards in the use of classifications whilst taking the local context and the particular history of racial terminology in South Africa into account.

The last contribution I will discuss, which is also the final chapter of the book, brings the collection to a close in answering the central question of the book: the relationship between race and research ethics. That question is whether or not the use of race in research is ethically justified. This is the topic of Keymanthri Moodley's contribution, 'Science, Race and Ethics'.[13]

In the context of research ethics, justification must incorporate both social and scientific reasons that show the benefits of undertaking research in one particular fashion rather than another. Ethical research, Moodley suggests, weighs the relative costs and benefits of research. The eight criteria Moodley offers for undertaking ethical research are meant to help decision-makers make a cost–benefit assessment. These eight criteria are: (1) establishing collaborative partnerships, (2) considering the social value of a study, (3) ensuring its scientific validity, (4) ensuring the fair selection of participants, (5) assessing the risk–benefit ratio of a study, (6) going through an independent ethics review, (7) having informed consent, and (8) ensuring respect for participants. These are practical criteria that can be used as a guideline in considering the ethics of a study.

Moodley's chapter begins with an overview of instances of scientific malpractice and unethical scientific experimentation on human subjects. Local examples she touches on are the infamous experiments of South Africa's Project Coast led by Wouter Basson to the Bezwoda case, and some of her international examples are the eugenics programmes of the 20th century and the US Tuskegee syphilis study. Different lessons can be drawn from the lack of research ethics in each case, specifically in respect to the blatant disregard of the human rights of the people experimented on in each of these cases. The cases explored are woven together by their racist ideological underpinnings that enabled justifications for targeting particular demographics for unethical treatment in scientific settings. The scientific merits for undertaking many of these studies were in themselves fundamentally questionable beyond their more general ethical failings.

As a precondition for research being ethical, Moodley posits that the research must be scientifically sound. As the methodology of the Nieuwoudt study was not scientifically sound, it should not have passed ethics review (note: the protocol used in the Nieuwoudt study was found not to match the protocol approved by the REC).[4] It is under such a framing that Moodley says 'this study is regarded as unscientific and therefore unethical'[13]. Her point is that expertise is central to the ability of researchers to be able to carry out ethical research. In this regard, the question arises as to why cognitive function was a topic seen fit to be explored by sport scientists as well as if any of them had the relevant expertise to undertake such a study. Addressing these questions may have led the researchers to make better decisions about this study, if to undertake it at all.

Researchers may not always have a fair idea of their own abilities or limitations. This is where the role of peer review and academic checks and balances have a role to play. Checks and balances present an opportunity to interrogate the assumptions researchers may have and

to test the social and scientific viability of their studies. This is the role of a REC, and it is critically important when considering difficult topics. In this regard, Moodley suggests that:

> *Given our history, any study that attempts to link race with cognitive function must be carefully considered by a REC. This is why the National Health Research Ethics Council (NHREC) in South Africa has guidelines for who should serve on a REC. Members should represent the demographic profile of the country to ensure sensitivity to local context in order for a proper risk assessment to be made.*[13]

This is perhaps challenging to do at universities like Stellenbosch University which remain predominantly white in both academic staff and student demographics. Because many South African universities remain untransformed in similar ways, universities struggle to meet these conditions. Thus, oversights on certain issues may be a direct result of these more institutional issues and the institutional culture under which research is assessed and produced. The institutional orientation and demographics of the University itself may be structurally biased towards the poor outcomes manifest in caricature in the Nieuwoudt paper. Moving forward, my hope is that the Nieuwoudt paper could bring urgency to calls for transformation and spur on efforts to undertake some fundamental change in our universities' structural orientations, pedagogy and research traditions.

An important lesson to come out of the Nieuwoudt experience that has been somewhat overlooked is the importance of research ethics for researchers themselves. If the details of the Nieuwoudt study are significantly different from what was approved by the REC,[4] then the researchers missed the opportunity to subject their work to quality control checks from which they could have benefitted. Nevertheless, as many of the contributions in *Fault Lines* suggest, the problems of race and research run far deeper than the operation of RECs. So long as those problems are not resolved, we can expect the production of racist research to be recurrent.

References

1. Nieuwoudt S, Dickie KE, Coetsee C, Engelbrecht L, Terblanche E. [Retracted article] Age- and education-related effects on cognitive functioning in Colored South African women. Aging Neuropsychol Cogn. 2020;27(3):321–337. https://doi.org/10.1080/13825585.2019.1598538
2. PsySSA. An open critique of the Nieuwoudt and colleagues (2019) study on Coloured women by the Division for Research and Methodology on behalf of the Psychological Society of South Africa [document on the Internet]. c2020 [cited 2020 Jun 04]. Available from: https://www.psyssa.com/wp-content/uploads/2019/04/PsySSA-Statement.pdf
3. Hendricks L, Kramer S, Ratele R. Research shouldn't be a dirty thought, but race is a problematic construct. S Afr J Psychol. 2019;49(3):308–311. https://doi.org/10.1177/0081246319852548
4. Cloete E. Update on outcome of formal investigation: "The Article" [webpage on the Internet]. c2020 [cited 2020 Jun 12]. Available from: http://www.sun.ac.za/english/Lists/news/DispForm.aspx?ID=7426
5. Editors and Publisher of *Aging, Neuropsychology, and Cognition*. Statement of retraction: Age- and education-related effects on cognitive functioning in Colored South African women. Aging Neuropsychol Cogn. 2019. https://doi.org/10.1080/13825585.2019.1614759
6. Jansen J, Walters C, editors. Fault Lines: A primer on race, science and society. Stellenbosch: SunMedia; 2020.
7. Jansen J. Introduction. In: Jansen J, Walters C, editors. Fault Lines: A primer on race, science and society. Stellenbosch: SunMedia; 2020. p. 10.
8. Bardien-Kruger S, Müller-Nedebock A. The role of genetics in racial categorisation of humans. In: Jansen J, Walters C, editors. Fault lines: A primer on race, science and society. Stellenbosch: SunMedia; 2020. p. 21–39.
9. Baker JL, Rotimi CN, Shriner D. Human ancestry correlates with language and reveals that race is not an objective genomic classifier. Sci Rep. 2017;7, Art. #1572, 10 pages. https://doi.org/10.1038/s41598-017-01837-7
10. Maglo KN, Mersha TB, Martin LJ. Population genomics and the statistical values of race: An interdisciplinary perspective on the biological classification of human populations and implications for clinical genetic epidemiological research. Front Genet. 2016;7(22):1–13.
11. Msimang PM. Social races in biological settings. In: Lorusso L, Winther RG, editors. Remapping race. Routledge; in press 2021.
12. Volmink J, Hendricks L, Mazibuko L, Swartz L. Race and health: Dilemmas of the South African health researcher. In: Jansen J, Walters C, editors. Fault lines: A primer on race, science and society. Stellenbosch: SunMedia; 2020. p. 122.
13. Moodley K. Science, race and ethics. In: Jansen J, Walters C, editors. Fault lines: A primer on race, science and society. Stellenbosch: SunMedia; 2020. p. 253–273.

AUTHOR:
Chenjerayi Kashangura[1]

AFFILIATION:
[1]Biological Sciences Department, University of Zimbabwe, Harare, Zimbabwe

CORRESPONDENCE TO:
Chenjerayi Kashangura

EMAIL:
ckashangura@kutsaga.co.zw

HOW TO CITE:
Kashangura C. Enhanced education on vaccines can reduce the scourge of vaccine rejection and hesitation. S Afr J Sci. 2020;116(9/10), Art. #8143, 2 pages. https://doi.org/10.17159/sajs.2020/8143

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
artificial intelligence, bioengineering, deep learning, virus, virtual reality, gaming

PUBLISHED:
29 September 2020

Enhanced education on vaccines can reduce the scourge of vaccine rejection and hesitation

Development and improvement of vaccines are benefitting from advances in life sciences fields such as synthetic biology. However, amidst this technological revolution, the field of vaccinology faces challenges in keeping up with evolving viral pathogens and adverse vaccine response outcomes due to human genetic diversity. Moreover, anti-vaccination attitudes create a higher incidence of vaccine hesitation and rejection.

Global health is facing challenges from evolving existing pathogens that are becoming multidrug resistant and thus reducing the available therapeutic interventions that can be employed. Moreover, novel pathogens are emerging, such as the current pandemic of SARS-Coronavirus 2 (SARS–Cov-2). Some of the hot topics in vaccination – namely efficacy, impact and attitudes – were highlighted in an Elsevier Vaccine Special Article Collection[1] that also identified some of the challenges in vaccinology. These challenges include anti-vaccine thinking that leads to increased levels of vaccine rejection and the need to work fast to produce a vaccine such as the annual flu vaccine which must match new strains.[1] Developing vaccines is usually a race against time as the viral particles mutate and recombine and thus require annual development *ab initio*. Successful vaccinology depends partly on the vaccines interacting with the immune system and producing an immune response similar to that produced by the natural infection, but usually not creating clinical disease symptoms and potential complications.[2] Vaccines can be in the form of live attenuated pathogen, virus-like particles, killed sub-unit, killed virus and viral components such as polysaccharide and protein[3] that result in long-term protection that requires the persistence of vaccine antibodies and/or the generation of immune memory cells that are capable of rapid and effective reactivation upon subsequent exposure to the pathogen.[3]

Genetic diversity in populations is essential as high levels of diversity usually correspond to fitness or adaptability to environments. Thus, any change in the environment will most likely have a corresponding adaptive genome already present in the population. A population that is highly genetically uniform would need mutations to occur to adapt to a new environment and, in the absence of mutations, fitness is lowered, and the population tends towards extinction. However, for vaccinology, high genetic diversity is not ideal as the interaction of genetic factors, such as polymorphisms in genes encoding immune response proteins – HLA molecules, cytokines and cytokine receptors – and environmental factors such as the dose, route of administration and quality of antigen can result in non-uniform or unexpected outcomes.[4,5] These may include adverse side effects or complications after the vaccine is administered. This genetic variation in both humans and the pathogen causes antigenic variation in the infectious agent and a high inter-individual variability in the human response to the vaccine[6] which prevents the attainment of a universally effective vaccine. On the other hand, a genetically uniform population would potentially have uniform outcomes after administering a vaccine. Africa has rich human genetic diversity[7] which needs to be taken into account in developing and evaluating vaccines.

Enhanced vaccine development in the form of multidisciplinary synergistic approaches may enable novel ways to develop vaccines that utilise the rapidly growing field of artificial intelligence (AI) *in silico* deep learning, deep reasoning, reading and vision capabilities. Novel algorithms can reduce genetic diversity to uniform demes, predict[8], model and select the strain to use. In addition, the algorithms can be developed to: consider human genetic variants that determine disease susceptibility[9], identify potential vaccine-adverse outcomes, genetic signatures, suggest improvements to a candidate therapeutic agent and take into consideration the existing genetic diversity of the human species races, local demes and that within the pathogen. This would assist in producing a situation of high human genetic diversity within the population separated *in silico* into genetically uniform demes and the prediction of potential vaccine-adverse outcomes signatures. Vaccines for genetically uniform populations would be likely to reduce adverse side effects. Building on the experience and outcomes of the first phase of algorithm development and testing, AI algorithms that predict the ability of potential vaccine candidates to 'inhibit pathogen multiplication' or 'cell receptor binding' might be developed. Testing algorithms using available molecular databases of protein and nucleic acid sequences of components of pathogen particles, pathogen therapeutic agents, and available literature may enable rapid discovery or suggest the form of a candidate therapeutic agent. Employing AI in vaccine development has potential in determining *in silico* whether one or several different vaccines are required against a pathogen. The efficacy of the candidate therapeutic agent can then be improved through bioengineering, for example, synthetically derived nucleic acid or protein domains that are tailor-made for the identified different genetic groups that elicit an immune response in a safe way that does not produce clinical symptoms.

However, the fast production of an effective vaccine does not ensure vaccine success, as anti-vaccine beliefs are becoming widespread.[1] Thus, educating the public about vaccines, both novel and in use, needs to move from disseminating vaccine facts to 'enhanced vaccine education'.

Social media is one route through which polarised attitudes towards vaccination are popularised.[10] Enhanced vaccine education may take a multidisciplinary approach that includes traditionally non-allied activities such as virtual reality games, smartphone apps and cartoons, to conceptualise vaccinology processes and their attendant challenges. Teams can develop games and apps that present vaccine development, evaluation and vaccine facts in 'fun' ways to the public; these can be made available on online app stores, on social media platforms, in children's gaming arcades and even as part of the school curriculum. The educational media can cater for different age groups. This approach can potentially counter the 'infodemic' on vaccine science and reduce incidences of vaccine hesitation and rejection.

Vaccine development and education can also benefit from a further paradigm shift towards development specific to localised genetic diversity and other ways to develop antivirals and thus partly address the current challenges that lead to vaccine hesitation and rejection.

References

1. Elsevier Vaccine Special Article Collection. Hot topics in vaccination: Efficacy, impact and attitudes. Vaccine. 2018;36 [cited 2020 Mar 18]. Available from: https://www.elsevier.com/life-sciences/journals/hot-topics-in-vaccination
2. US Centers for Disease Control and Prevention (CDC). Principles of vaccination. In: Epidemiology and prevention of vaccine-preventable diseases. 13th ed. Washington DC: Public Health Foundation, CDC; 2015. p. 1–8. Available from: https://www.cdc.gov/vaccines/pubs/pinkbook/prinvac.html
3. Siegrist CA. Vaccine immunology. In: Vaccines. 6th ed. Philadelphia, PA: Elsevier Saunders; 2013. p. 14–32. https://doi.org/10.1016/B978-1-4557-0090-5.00004-5
4. Newport MJ. The genetic regulation of infant immune responses to vaccination. Front Immunol. 2015;6, Art. #18, 5 pages. https://doi.org/10.3389/fimmu.2015.00018
5. O'Connor D, Png E, Khor CC, Snape MD, Hill AVS, Van der Klis M, et al. Common genetic variations associated with the persistence of immunity following childhood immunisation. Cell Rep. 2019;27(11):3241–3253. https://doi.org/10.1016/j.celrep.2019.05.053
6. Lemaire D, Barbosa T, Rihet P. Coping with genetic diversity: The contribution of pathogen and human genomics to modern vaccinology. Braz J Med Biol Res. 2012;45(5):376–385. https://doi.org/10.1590/s0100-879x2011007500142
7. Sirugo G, Hennig BJ, Adeyemo AA, Matimba A, Newport MJ, Ibrahim ME, et al. Genetic studies of African populations: An overview on disease susceptibility and response to vaccines and therapeutics. Hum Genet. 2008;123(6):557–598. https://doi.org/10.1007/s00439-008-0511-y
8. Stokes JM, Yang K, Swanson K, Jin W, Cubillos-Ruiz A, Donghia NM, et al. A deep learning approach to antibiotic discovery. Cell. 2020;180(4):688–702. https://doi.org/10.1016/j.cell.2020.01.021
9. Hill AVS. Genetics and genomics of infectious disease susceptibility. Br Med Bull. 1995;55(2):401–413. https://doi.org/10.1258/0007142991902457
10. Gesualdo F, Zamperini N, Tozzi AE. To talk better about vaccines, we should talk less about vaccines. Vaccine. 2018;36(34):5107–5108. https://doi.org/10.1016/j.vaccine.2018.07.025

AUTHOR:
Steven D. Johnson[1]

AFFILIATION:
[1]Centre for Functional Biodiversity, School of Life Sciences, University of KwaZulu-Natal, Pietermaritzburg, South Africa

CORRESPONDENCE TO:
Steven Johnson

EMAIL:
johnsonSD@ukzn.ac.za

DATES:
Received: 27 July 2020
Revised: 07 Aug. 2020
Accepted: 07 Aug. 2020
Published: 21 Sep. 2020

HOW TO CITE:
Johnson SD. Peer review versus the *h*-index for evaluation of individual researchers in the biological sciences. S Afr J Sci. 2020;116(9/10), Art. #8700, 5 pages. https://doi.org/10.17159/sajs.2020/8700

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:
bibliometrics, Hirsch's *h*-index, National Research Foundation, NRF rating, research administration

FUNDING:
South African National Research Foundation (grant 46372)

Peer review versus the *h*-index for evaluation of individual researchers in the biological sciences

Past performance is a key consideration when rationalising the allocation of grants and other opportunities to individual researchers. The National Research Foundation of South Africa (NRF) has long used a highly structured system of 'rating' the past performance of individual researchers. This system relies heavily on peer review, and has seldom been benchmarked against bibliometric measures of research performance such as Hirsch's *h*-index. Here I use data for about 600 rated researchers in the biological sciences to evaluate the extent to which outcomes of peer review correspond to bibliometric measures of research performance. The analysis revealed that values of the *h*-index based on the Scopus database are typically 5–20 for researchers placed in the NRF's C rating category ('established'), 20–40 for those in the B rating category ('considerable international recognition') and >40 for those in the A rating category ('leading international scholars'). Despite concerns that citation patterns differ among disciplines, the mean *h*-index per rating category was remarkably consistent across five different disciplines in the biological sciences, namely animal sciences, plant sciences, ecology, microbiology and biochemistry/genetics. This observation suggests that the NRF rating system is equitable in the sense that the outcomes of peer review are generally consistent with bibliometric measures of research performance across different disciplines in the biological sciences. However, the study did reveal some notable discrepancies which could reflect either bias in the peer-review process or shortcomings in the bibliometric measures, or both.

Significance:

- NRF rating categorisations (estimates of standing in a research field based on peer review) are a reasonably good predictor of the *h*-index of individual researchers in the biological sciences.
- The relationships between rating categorisations and the *h*-index are remarkably consistent across five sub-disciplines in the biological sciences.
- Peer review and the *h*-index were compared in terms of their relative advantages and disadvantages and the combined use of both approaches is advocated for measurement of research performance.

Introduction

Peer review is one of the pillars of the global research enterprise.[1] At a political level, national governments often steer overall research activities in particular directions through funding allocations to programmes, but they also almost always devolve the final allocation of resources to a system of peer review. The logic is that peers are considered to be in the best position to evaluate the quality of past and proposed research, even though it is also acknowledged that reviewers can exhibit bias, which can be either explicit, such as overt competition among researchers, or implicit, such as underlying prejudice according to race, gender and the perceived status of the institution to which a researcher is affiliated.[1] The peer-review system is under enormous strain worldwide and grant administrators increasingly struggle to obtain quality reviews of funding proposals.[2] Peer review, whether of grants or submitted manuscripts, is usually performed without added remuneration and there are thus limits to the time that researchers are willing to allocate to this process, particularly if it involves an assessment of the entire track record of each individual researcher.

One possible way of reducing the burden on peer reviewers (and the institutions that manage the process) is to separate periodic evaluation of the track record of the applicant from the evaluation of proposed research, such that a single measure of the applicant's track record can be used for multiple decision-making processes. This approach has been adopted in South Africa for several decades and takes the form of peer evaluation of the research performance of researchers in the higher education and research sector. This evaluation is usually based on the opinions of about six peers (some suggested by the candidate and some suggested by a panel of discipline-specific experts) and is codified as a particular 'rating' which lasts for a period of 6 years. This information can then be used by administrative panels when making grant allocations or for other purposes such as informing a university selection committee about the past performance and general standing of an applicant in a research field.

The availability of massive computerised databases of publications that include citation information has led to the development of bibliometric measures of research performance. By far the most widely adopted of these is Hirsch's *h*-index, which is simply the number *h* of papers that have been cited at least *h* times.[3] The *h*-index is intended to strike a balance between the total number of citations, which may be unduly influenced by a few very well cited papers, and the total number of publications, which may not reflect the actual impact of the research in terms of citations.[4] The *h*-index has been shown to be closely associated with measures of academic standing in a field.[5,6] Potential drawbacks of the *h*-index include its insensitivity to number of authors, author positions, and discipline-specific citation patterns.[4,7-11] Like the metric of total citations, the *h*-index also shows a ratchet effect whereby it will continue to increase even after a researcher has become inactive.[4]

There have been very few attempts worldwide to determine the extent of agreement between peer review and bibliometric measures. Most of these involve studies of whether the *h*-index can predict the outcomes of applications for fellowships[12,13] or future career trajectories[14]. Hirsch[9] has characterised the *h*-index as 'an indicator

of the impact of a researcher on the development of his or her scientific field'. This is uncannily similar to the objectives of the National Research Foundation (NRF) rating system in South Africa. Studies of correlations between peer assessment of research standing, such as NRF ratings, and the *h*-index are particularly valuable, but remain rare.[5,8,15,16] Lovegrove and Johnson[17] analysed data for a small sample (163) of botanists and zoologists in South Africa and found that the outcome of peer review of research standing in the form of NRF rating categories was fairly well correlated with the ISI-based *h*-index. They did not attempt to analyse trends across other disciplines in biology, however, and the ranges and means they obtained for the *h*-index in each rating category need to be updated, given the large increase in publications worldwide over the past 12 years.

The aim of the present study was to establish, across different disciplines of biology, the relations between estimates of the standing of researchers in their field based on peer review and those based on bibliometric analysis of their *h*-index.

Methods

I used the public database of NRF ratings assigned to 4176 South African researchers, available at www.nrf.ac.za, which was last updated on 30 June 2020. I filtered this database down to 644 researchers who mentioned 'Biological Sciences' as one of their primary disciplines (Supplementary table 1). I was able to find the Scopus-based *h*-index values for 614 of these researchers (searches took place 18–19 July 2020). Values of the Scopus-based *h*-index are generally very similar to those based on the Web of Science.[18] I did not use Google Scholar because Google Scholar profiles that are not frequently curated often include papers that are not authored by the researcher, thus inflating their actual *h*-index.

I was able to assign 569 researchers to a sub-discipline, usually based on their own statement of a sub-discipline in their NRF profile, or, more rarely, by examining the content of papers in their Scopus profile or consulting with other researchers in the sub-discipline. I used the following sub-discipline categories (number of researchers): 'Ecology' (137), 'Plant Sciences' (75), 'Animal Sciences' (127), 'Microbiology' (115) and 'Biochemistry, Molecular Biology and Genetics' (115). Palaeontologists were allocated to the Animal Sciences category only if their main focus is on animal structures and evolution rather than geology. These disciplines are based loosely on those used by Scopus and do not correspond exactly to the specialist committees used by NRF to oversee the results of peer review. For example, I have grouped all ecologists into one category, regardless of whether they would fall under the plant sciences or animal sciences committees.

I analysed values of the *h*-index of researchers to determine the degree to which these could be predicted by peer-review outcomes (the rating categories in which these researchers had been placed) as well as according to the specific discipline of the researchers. The three rating categories used by the NRF that were analysed (and the number of researchers in these) were A ('leading international scholars', $n=27$), B ('scholars with considerable international recognition', $n=129$) and C ('established scholars', $n=414$). I did not analyse rating sub-categories, e.g. C1, C2 and C3, as this information is not made public by the NRF. Even if the rating sub-category data could be obtained on the condition of confidentiality, as was the case in the analysis by Lovegrove and Johnson[17], analysing them here without making the data available as an appendix would violate the principle of data transparency. I did not analyse researchers who applied for a rating and were unsuccessful as this information is not publicly available; nor did I analyse the Y and P rating categories as these specifically apply to early-career researchers who are evaluated largely according to future potential and not on their past track record in terms of impact within their fields.

Values of the *h*-index were skewed with a longer tail to the right and were thus analysed using a generalised linear model which incorporated a gamma distribution and log link function (implemented in SPSS 25, IBM Corp.). Rating category, sub-discipline and their interaction were treated as fixed predictors, and time (0–5 years) since the rating was awarded was treated as a covariate. Model significance was assessed using likelihood ratios and post-hoc comparisons among means were based on the Dunn–Šidák method. In statistical terms, this model tests the null hypothesis that different rating category allocations are drawn from the same distribution of the *h*-index. An equally valid approach not used here would be to use a logistic model to assess whether the *h*-index is a good predictor of the probability of different rating categorisations.[8]

Results

The distributions of raw (non-adjusted) *h*-indices according to rating category revealed clear clustering, but with some overlap of values between categories (Figure 1). The *h*-index values for researchers differed significantly according to rating category ($\chi^2 = 399.9$, $p<0.0001$), sub-discipline ($\chi^2 = 14.2$, $p=0.006$) and time since rating ($\chi^2 = 399.9$, $p<0.0001$), but not by the interaction between rating category and discipline ($\chi^2 = 10.4$, $p=0.234$).

The overall model-adjusted means (±s.e.) for the *h*-index were 16.4±0.29 for researchers in the C category, 29.8±0.96 for researchers in the B category and 57.4±4.16 for researchers in the A category (all category means differed significantly, post-hoc tests $p<0.0001$). The ranges in mean *h*-index across disciplines were 15.4–18.0 for the C category, 25.8–31.8 for the B category and 38.0–81.5 for the A category (Table 1).

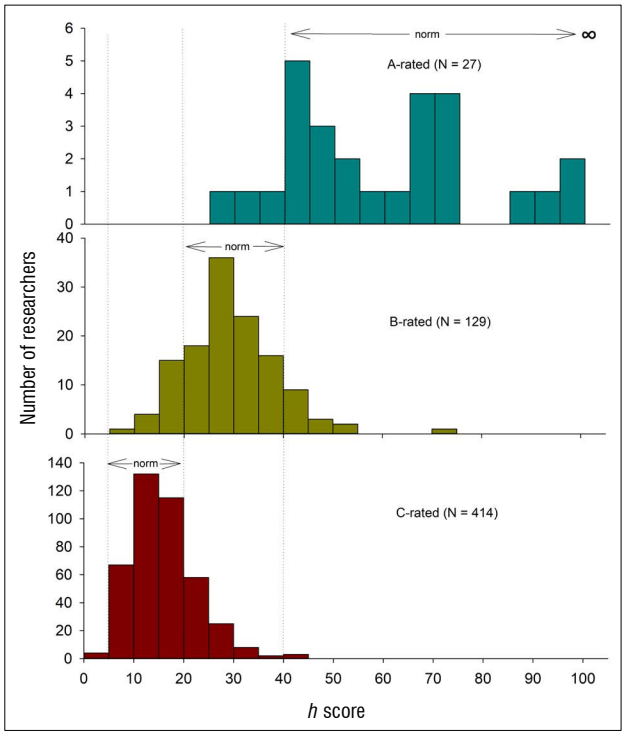


Figure 1: The distribution of values of Hirsch's *h*-index for researchers in the biological sciences placed in the three main rating categories used by the National Research Foundation of South Africa: A = 'leading international scholars', B = 'scholars with considerable international recognition' and C = 'established scholars'.

Discussion

These results indicate that there is substantial agreement between the evaluation of research performance based on peer review and values of the *h*-index based on bibliometrics across a wide range of sub-disciplines in biology (Table 1). The consistency across disciplines is particularly evident for the B and C rating categories (Table 1). The greatest variation in the mean *h*-index across disciplines, notably a twofold difference between animal sciences and molecular biology, is found in the A category (Table 1), but as this category has no upper bound and includes a relatively small number of researchers, the mean value for a sub-discipline can be shifted upwards considerably by one or two researchers with particularly high values of the *h*-index.

Finding that the *h*-index is largely congruent with NRF ratings does not automatically imply that measures of research performance based on peer review can simply be supplanted by bibliometrics. The reason for this caution is that both approaches have disadvantages, and a combination of both may offer the best safeguard against unfair evaluation of an individual. Given that the advantages and disadvantages of peer review, partially summarised in Table 2, are well known[1,15], I focus here on the advantages and risks of relying on the *h*-index as a measure of research performance.

The purpose of measuring research performance is usually to make decisions about allocations of public funds in the form of grants. It is therefore important that the measure reflects overall research competence and ability. In this sense, the *h*-index has some serious drawbacks, such as favouring authors who publish as part of large teams, including consortia.[7,19] For example, an author who is a minor (middle) author of a large number of multi-authored papers may quickly develop a healthy *h*-index and yet may not have the requisite experience in managing research or writing papers.[9,20] This issue would probably be flagged by peer reviewers. There is an increasing trend for papers to have large numbers of authors. The charitable view is that this reflects a genuine increase in collaboration among researchers, but a more cynical view is that it reflects a form of collaborative gaming of the system by authors who wish to collectively increase their numbers of publications and citations.[21] A curious footnote to this issue is that in South Africa, government incentive funding for each publication is divided among authors, thus providing a perverse incentive (counter to that applied by the *h*-index) for researchers to minimise the number of co-authors. Another form of gaming by authors is to focus on writing review articles, rather than to conduct original laboratory or field-based research, simply because review articles are well known to garner more citations.[21] The need to more fairly reward authors for original research has recently been recognised in the 'San Francisco Declaration on Research Assessment' (https://sfdora.org/) which includes an advisory for authors to cite original research rather than reviews wherever it is possible to do so.

The *h*-index ratchets upwards throughout a person's career and even continues to increase after they have become research-inactive. This is potentially a serious problem given that the purpose of performance evaluation is usually to allocate resources to individuals who are currently active. Reviewers for the NRF rating system in South Africa, for example, are expected to focus on the quality and impact of the research performed in the 8 years preceding the evaluation, and the overall *h*-index can be misleading in this regard. Only by focusing on the year-by-year trends in citations and publication quantity and quality over that period can a panel gain insight into the recent impact of a person in their field. Of course, the situation is completely different if the purpose of using the *h*-index is to award a prize for career impact to someone who has retired or who is approaching retirement, as opposed to using the information for allocation of resources. It also seems obvious that the *h*-index has limited utility for evaluating early-career researchers, although there have been some studies that have shown that early-career values of the *h*-index (medians around 2–3) do correlate to some extent with the peer-review outcomes of fellowship applications.[13] Personally I would be sceptical of relying too heavily on the *h*-index to evaluate early-career researchers for purposes such as awarding postdoctoral grants as it may favour applicants in a manner that is directly proportional to the number of years since completing (or commencing) their PhD, even if those intervening years were not particularly productive.

Given these drawbacks of the *h*-index, should it be used in performance evaluation at all? The answer, I believe, is an emphatic 'yes'. If the evaluation of a researcher by a peer-review process leads to an outcome that is incongruent with their *h*-index, then a panel needs to consider the basis for the incongruence. For example, a person with an *h*-index of 10 whose peers write reports that place them in the B category ('considerable international recognition') may have benefitted from reviewers who had been primed to provide favourable reports, and the case should be reconsidered. It would be very difficult to detect this misfeasance without the use of a benchmark provided by the *h*-index. However, it can also be the case that peer review is sometimes the more reliable measure. For example, a person whose peers write reports suggestive of a B rating may have an *h*-index of 60. Further investigation by a panel may reveal that such a person is no longer fully research-active, and the final decision may be more in accordance with the peer review than the *h*-index. It is even possible that the peer reviewers themselves make use of the *h*-index when asked to evaluate a candidate for rating, especially one with whom they are not very familiar.[21] Thus it could be the case that measures of research standing in a field based on peer review, such as NRF ratings, are, in fact, already being informed by the *h*-index. This lack

Table 1: Model-adjusted (marginal) mean *h*-index values for sub-disciplines within each NRF rating category

Rating	Discipline	Researchers	Mean	s.e.	95% Confidence interval	
					Lower	Upper
A	Animal Sciences	4	38.0	6.82	26.69	53.98
	Biochemistry, Molecular Biology, Genetics	4	81.6	14.65	57.32	115.95
	Ecology	6	64.5	9.47	48.40	86.03
	Microbiology	9	59.7	7.15	47.18	75.46
	Plant Sciences	4	52.5	9.48	36.83	74.77
B	Animal Sciences	31	25.8	1.66	22.69	29.24
	Biochemistry, Molecular Biology, Genetics	24	29.3	2.16	25.41	33.89
	Ecology	34	30.6	1.89	27.15	34.59
	Microbiology	21	31.8	2.50	27.22	37.07
	Plant Sciences	19	31.7	2.62	27.00	37.31
C	Animal Sciences	92	15.4	0.58	14.26	16.52
	Biochemistry, Molecular Biology, Genetics	87	16.4	0.63	15.19	17.67
	Ecology	98	18.0	0.66	16.79	19.38
	Microbiology	85	16.9	0.66	15.62	18.21
	Plant Sciences	52	15.7	0.78	14.20	17.27

Table 2: A comparison of peer-review and the *h*-index in terms of advantages and disadvantages

Peer review		*h*-index	
Advantages	**Disadvantages**	**Advantages**	**Disadvantages**
• Can evaluate recent research performance • Can account for complexities such as author position and the contribution of an individual to a multi-authored paper	• Reviewer bias • Reviewer fatigue • Administratively complex • Requires textual interpretation by panel members	• Not sensitive to citations of single papers • Considers impact of papers, not journal impact factors • Inexpensive to administer	• Does not decrease when research activity slows (ratchet effect) • Does not take number of co-authors into account • Does not take author positions into account • Not suitable for evaluating early-career researchers • Can be 'gamed' by self-citation of papers just below *h*-index threshold • Relatively insensitive to massively cited 'big hit' papers (can also be interpreted as an advantage) • May vary among disciplines

of clear separation between independent and dependent variables means that any statistical test of associations between the *h*-index and ratings of research standing, such as those in the present study, should be viewed as no more than crude heuristic approximations. This also means that any characterisation of peer review as 'subjective' and the *h*-index as 'objective' would be overly simplistic.[22] It could even be argued, quite reasonably, that the *h*-index itself is already a form of peer review as most citations are essentially a form of validation by peers – although it is of course possible for papers to be cited as examples of faulty research.

An unexpected result of this analysis was the huge inflation (roughly doubling) of *h*-index values within each rating category in the biological sciences since the previous analyses by Lovegrove and Johnson[17] and Fedderke[8]. For example, the lower *h*-index threshold for a B rating in the 2008 study was about 10 and is now about 20, and for an A rating this lower threshold has shifted from about 20 to 40 (Figure 1). It is not clear why this is the case, but one possibility is that the yearly increase in publications is outstripping the number of new researchers in the global system, thus leading to increasing values of the *h*-index. Another (related) possibility is that the number of authors per paper is increasing, leading to more papers for the same researcher effort. The most cynical and depressing explanation is that there is limited turnover of new researchers, particularly in the A and B categories, and so the NRF system is simply repeatedly re-evaluating much the same cohort of people whose *h*-index is increasing with time. Whatever the reason, it is clear that the *h*-index norms for each rating category are moving targets and, for administrative purposes, may need to be adjusted upwards every few years.

As fitted by eye, the *h*-index norms in the data for researchers in the biological sciences appear to be about 5–20 for the C category, 20–40 for the B category and >40 for the A category (Figure 1). These norms encompass 75.8% of researchers in the C category, 69.7% of researchers in the B category and 88.8% of researchers in the A category. In this analysis, there were 142 researchers (c. 25% of the total) whose ratings lie outside these norms and some of these represent marked outliers (Figure 1). These outliers could reflect failures of the peer-review system, such as review reports that are negatively biased or primed to be favourable when they should not be, or they could represent special cases fairly deliberated on by a panel, such as cases of researchers who have high a *h*-index, but whose research standing has waned or whose contributions to well-cited papers were relatively minor.

It should be noted here that the average gap between rating and recording of the *h*-index in this data set is c. 2.5 years, meaning that the actual *h*-index at the time of rating would have typically been slightly lower for each researcher. I controlled for the time gap between rating and scoring of each *h*-index in the statistical model by including it as a covariate. Although the marginal means for each category are adjusted according to this covariate, the mean *h*-index at the time that rating took place will still be slightly overestimated for some researchers. This problem of overestimation of the *h*-index at the actual time of rating also applies to previous analyses[8,17] and cannot account for the above-mentioned inflation of the *h*-index per rating category over the past decade and more.

It is uncertain whether the NRF will continue with the rating of researchers, even though it has not publicly indicated any intention to phase out the system. It has already been pointed out that the rating system seems to be quite disconnected with allocations of large state resources to the South African Research Chairs initiative[8] and rating itself is no longer connected with any significant 'incentive' funding allocations to researchers.[23] On the basis of cost and effort relative to application, it seems unlikely that the rating system will persist, but the general need for an evaluation of the standing of a researcher in a discipline will remain relevant even if the process is eventually bundled into the evaluation of grant applications.

It would be straightforward to extend this study to other disciplines besides the biological sciences, as was done previously by Fedderke[8] who used data derived from Google Scholar and found strong discipline-specific associations between the *h*-index and probabilities of ratings. Such an exercise would further clarify whether *h*-index norms are more or less universal across the natural and social sciences, or whether each broad discipline needs a different set of norms with which to benchmark the assignment of researchers to different rating categories. There is no doubt that the *h*-index can serve as a valuable benchmark when peer review is biased or blatantly unfair. Some authors have even expressed the view that a rating system based on bibliometrics would be viewed as being more progressive and transparent.[8,24] In cases where the spread of data such as those in Figure 1 are available, I would even venture to suggest that under conditions of severe resource or time constraints, organisations such as the NRF could consider using the *h*-index in a discipline-adjusted manner as the sole measure of the standing of a researcher in their field. However, peer review will usually provide more nuanced interpretations of career impact than will a single number based on bibliometrics, and should be used whenever it is feasible. For early-career (<8 years) researchers, whose *h*-index will almost always be modest, my advice would be to simply ignore this whole debate and focus on publishing captivating high-quality papers, because that remains the only way to make an impact on one's research field, no matter how it may ultimately be measured.

Acknowledgements

I thank Jane Carruthers for inviting this submission to SAJS as a contribution to international Peer Review Week. I am also grateful to Theresa Coetzer, Barry Lovegrove, Timo van der Niet, Daniel Johnson and Kathryn Johnson and two anonymous reviewers for comments on the manuscript, and Theresa Coetzer for assistance with the assigning of some of the researchers to sub-disciplines (the responsibility for any errors is mine). This research was supported by NRF grant 46372 as part of the DST-NRF South African Research Chairs Initiative.

Competing interests

I declare that there are no competing interests.

References

1. Smith R. Peer review: A flawed process at the heart of science and journals. J R Soc Med. 2006;99(4):178–182. https://doi.org/10.1177/014107680609900414
2. Nguyen VM, Haddaway NR, Gutowsky LFG, Wilson ADM, Gallagher AJ, Donaldson MR, et al. How long is too long in contemporary peer review? Perspectives from authors publishing in conservation biology journals. PLoS ONE. 2015;10(8), e0139783. https://doi.org/10.1371/journal.pone.0132557
3. Hirsch JE. An index to quantify an individual's scientific research output. Proc Natl Acad Sci USA. 2005;102(46):16569–16572. https://doi.org/10.1073/pnas.0507655102
4. Egghe L. The Hirsch index and related impact measures. Annu Rev Inform Sci Technol. 2010;44:65–114. https://doi.org/10.1002/aris.2010.1440440109
5. Benway BM, Kalidas P, Cabello JM, Bhayani SB. Does citation analysis reveal association between *h*-index and academic rank in urology? Urology. 2009;74(1):30–33. https://doi.org/10.1016/j.urology.2008.10.045
6. Van Raan AFJ. Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups. Scientometrics. 2006;67(3):491–502. https://doi.org/10.1556/Scient.67.2006.3.10
7. Barnes C. The *h*-index debate: An introduction for librarians. J Acad Libr. 2017;43(6):487–494. https://doi.org/10.1016/j.acalib.2017.08.013
8. Fedderke JW. The objectivity of National Research Foundation peer review in South Africa assessed against bibliometric indexes. Scientometrics. 2013;97(2):177–206. https://doi.org/10.1007/s11192-013-0981-0
9. Hirsch JE, Buela-Casal G. The meaning of the *h*-index. Int J Clin Health Psychol. 2014;14(2):161–164. https://doi.org/10.1016/S1697-2600(14)70050-X
10. Iglesias JE, Pecharroman C. Scaling the *h*-index for different scientific ISI fields. Scientometrics. 2007;73(3):303–320. https://doi.org/10.1007/s11192-007-1805-x
11. Kelly CD, Jennions MD. The *h* index and career assessment by numbers. Trends Ecol Evol. 2006;21(4):167–170. https://doi.org/10.1016/j.tree.2006.01.005
12. Bornmann L, Daniel HD. Does the *h*-index for ranking of scientists really work? Scientometrics. 2005;65(3):391–392. https://doi.org/10.1007/s11192-005-0281-4
13. Bornmann L, Wallon G, Ledin A. Does the committee peer review select the best applicants for funding? An investigation of the selection process for two European molecular biology organization programmes. PLoS ONE. 2008;3(10), e3480. https://doi.org/10.1371/journal.pone.0003480
14. Hirsch JE. Does the *h* index have predictive power? Proc Natl Acad Sci USA. 2007;104(49):19193–19198. https://doi.org/10.1073/pnas.0707962104
15. Derrick GE, Haynes A, Chapman S, Hall WD. The association between four citation metrics and peer rankings of research influence of Australian researchers in six fields of public health. PLoS ONE. 2011;6(4), e18521. https://doi.org/10.1371/journal.pone.0018521
16. Norris M, Oppenheim C. Peer review and the *h*-index: Two studies. J Informetr. 2010;4(3):221–232. https://doi.org/10.1016/j.joi.2009.11.001
17. Lovegrove BG, Johnson SD. Assessment of research performance in biology: How well do peer review and bibliometry correlate? Bioscience. 2008;58(2):160–164. https://doi.org/10.1641/B580210
18. Meho LI, Rogers Y. Citation counting, citation ranking, and *h*-index of human-computer interaction researchers: A comparison of Scopus and Web of Science. J Am Soc Inform Sci Technol. 2008;59(11):1711–1726. https://doi.org/10.1002/asi.20874
19. Wuchty S, Jones BF, Uzzi B. The increasing dominance of teams in production of knowledge. Science. 2007;316(5827):1036–1039. https://doi.org/10.1126/science.1136099
20. Stallings J, Vance E, Yang JS, Vannier MW, Liang JM, Pang LJ, et al. Determining scientific impact using a collaboration index. Proc Natl Acad Sci USA. 2013;110(24):9680–9685. https://doi.org/10.1073/pnas.1220184110
21. Chapman CA, Bicca-Marques JC, Calvignac-Spencer S, Fan PF, Fashing PJ, Gogarten J, et al. Games academics play and their consequences: How authorship, *h*-index and journal impact factors are shaping the future of academia. Proc Roy Soc B Biol Sci. 2019;286(1916), Art. #20192047. https://doi.org/10.1098/rspb.2019.2047
22. Costas R, Franssen T. Reflections around 'the cautionary use' of the *h*-index: Response to Teixeira da Silva and Dobranszki. Scientometrics. 2018;115(2):1125–1130. https://doi.org/10.1007/s11192-018-2683-0
23. Coldwell DA. A fair individualised university researcher rating system? A rejoinder to the current NRF debate. S Afr J Sci. 2019;115(3–4), Art. #5445. https://doi.org/10.17159/sajs.2019/5445
24. Callaghan C. A review of South Africa's National Research Foundation's ratings methodology from a social science perspective. S Afr J Sci. 2018;114(3–4), Art. #2017-0344. https://doi.org/10.17159/sajs.2018/20170344

AUTHORS:
Donrich Thaldar[1,2]
Marietjie Botes[1,2]
Bonginkosi Shozi[1,2]
Beverley Townsend[1,2]
Julian Kinderlerer[1,2]

AFFILIATIONS:
[1]School of Law, University of KwaZulu-Natal, Durban, South Africa
[2]African Health Research Flagship, University of KwaZulu-Natal, Durban, South Africa

CORRESPONDENCE TO:
Donrich Thaldar

EMAIL:
ThaldarD@ukzn.ac.za

DATES:
Received: 30 July 2019
Revised: 15 Apr. 2020
Accepted: 01 May 2020
Published: 29 Sep. 2020

HOW TO CITE:
Thaldar D, Botes M, Shozi B, Townsend B, Kinderlerer J. Human germline editing: Legal-ethical guidelines for South Africa. S Afr J Sci. 2020;116(9/10), Art. #6760, 7 pages. https://doi.org/10.17159/sajs.2020/6760

ARTICLE INCLUDES:
☒ Peer review
☐ Supplementary material

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

EDITOR:
Pascal Bessong

KEYWORDS:
genome editing, gene editing, CRISPR-Cas9, prospective person, health-care access, scientific freedom, assisted reproduction

FUNDING:
University of KwaZulu-Natal (African Health Research Flagship Grant); South African National Research Foundation (grant 116275)

Human germline editing: Legal-ethical guidelines for South Africa

Human germline editing holds much promise for improving people's lives, but at the same time this novel biotechnology raises ethical and legal questions. The South African ethics regulatory environment is problematic, as it prohibits all research on, and the clinical application of, human germline editing. By contrast, the South African legal regulatory environment allows a regulatory path that would, in principle, permit research on human germline editing. However, the legal regulation of the clinical application of human germline editing is uncertain. As such, the current ethical and legal positions in South Africa are in need of reform. Five guiding principles – aligned with the values of the Constitution – are proposed to guide ethical and legal policy reform regarding human germline editing in South Africa: (1) Given its potential to improve the lives of the people of South Africa, human germline editing should be regulated, not banned. (2) Human germline editing clinical applications should only be made accessible to the public if they are proven to be safe and effective. (3) Non-therapeutic human germline editing may be permissible, and should be regulated in the same way as therapeutic human germline editing. (4) The decision on whether to use germline gene editing on a prospective child, should, subject to Principle 2, be left to the prospective parents. (5) Concerns about exacerbating social inequalities should be addressed by measures to increase access. In conclusion, recommendations are made to policymakers and scientists contemplating research in this field.

Significance:

The ethical and legal positions regarding human germline editing in South Africa are comprehensively analysed. Furthermore, five guiding principles – aligned with the values of the Constitution – are proposed to guide much needed ethical and legal policy reform regarding human germline editing in South Africa.

Introduction

Human germline editing holds the promise of significant benefit, but also harbours secrets of yet unknown possible genetic risks and consequences for future generations.[1] These risks and benefits considered, the germline editing reportedly performed by Chinese scientist He Jiankui on human in vitro embryos, resulting in the birth of two girls with edited genomes[2], ignited a flurry of global legal-ethical debate on whether germline editing should be permissible, and if so, under what circumstances[3-7]. What is South Africa's current position in terms of legal and ethical regulation of human germline editing? And what *should* South Africa's current position be in light of the values of the South African Constitution? These are the questions that we grapple with in this article.

Terminology

Genome editing[8] (or gene editing) refers to the modification of the genome through targeted adding, replacing or removing one or more DNA base pairs in the genome, regardless of whether the modifications occur in a particular gene or a non-coding region of the genome[9]. Modern techniques used in genome editing are more precise than those that have in the past been used to genetically modify organisms, and include technologies such as CRISPR-CasX (where X is usually a digit e.g. 9), zinc finger nuclease (ZFN) transcription activator-like-effector based nucleases (TALEN) and meganucleases.[10,11] Genome editing can be performed in both somatic cells and germline cells.[9] Germline cells include early stage embryos, eggs, sperm and any cells that give rise to eggs and sperm.[1] Genome editing on germline cells (or germline editing in short) is usually more ethically and legally controversial, because the modification to the genome is intended to be heritable – in other words, the modification is likely to impact both the gene-edited individual and his or her genetic offspring. This article focuses exclusively on human germline editing.

Human germline editing is often divided into therapeutic and non-therapeutic germline editing. Typically, the term 'therapeutic' germline editing (or germline gene therapy) refers to the correction of a genetic defect in germ cells, with the aim of the gene-edited individual being born with a 'normal' genome; while the term 'non-therapeutic' germline editing refers to the modification of a normal genome in germ cells, with the aim of the gene-edited individual being born with an 'enhanced' genome.[12] Non-therapeutic germline editing therefore applies both to a modification that aims to bestow a health-related benefit to the gene-edited individual (such as immunity against contracting a certain illness) and to a modification that aims to bestow a non-health-related benefit to the gene-edited individual (such as higher intelligence – assuming it is possible).

For purposes of regulation, a further differentiation is relevant, namely between *research* on human germline editing and the *clinical application* of human germline editing. However, these terms, broadly understood, can overlap. With 'research' we refer to the production of generalisable knowledge about human germline editing through systematic investigation, including preclinical studies (such as in vitro human studies and animal trials)[13] and clinical trials. With 'clinical application' we refer to the use of modified germ cells in human reproduction. This will include clinical trials and – if shown to be safe and effective – the provision of human germline editing as a service to the public.

Analysis of the current position

Ethics guidelines

In this section, we analyse the relevant provisions of the three most prominent ethics instruments in South Africa, namely the ethics guidelines of the South African Department of Health[14], the Health Professions Council of South Africa (HPCSA)[15] and the South African Medical Research Council (MRC)[16].

The Department of Health's ethics guidelines have sections on genetics research and on genomics research that highlight general issues that are of ethical relevance with these types of research. However, the Department of Health's ethics guidelines are silent on the more specific topic of genome editing.

In 2008, the HPCSA published a code of ethical practice for medical biotechnology research in South Africa as Booklet 14 titled 'General ethical guidelines for biotechnology research in South Africa'.[15] Subsequently, and at the time of finalising this article, a different version of Booklet 14, dated 2016, was available on the HPCSA website. However, this version is a verbatim duplicate of Booklet 13, titled 'General ethical guidelines for health researchers', which provides only general guidance on health research. Given the duplication, we assume that this is an administrative error and so rather rely on the 2008 version of Booklet 14 that specifically provides guidance on germline gene therapy and research.

The HPCSA states (under heading '13.3 Gene Therapy Research') that 'no attempts should be made through the use of gene modification, to change human traits not associated with disease'. The HPCSA does not provide any indication as to its reasons for this position. The HPCSA further argues (under heading '13.3.2 Germ line gene therapy research') that because of the heritable nature of germline gene therapy, research relating to germline gene therapy is 'not acceptable'. Although the HPCSA does not define exactly what it means by 'therapy' (e.g. does it include health-related enhancements to the genome?), it appears from the reasoning (based on heritability) that the intention is to ban all forms of human germline editing research – and per implication the clinical application thereof. This we view as problematic. As we discuss below, heritability per se is certainly a factor to consider, but it is not necessarily a negative factor, or sufficient reason for a ban.

The ethics guidelines of the MRC appear to have contradictory positions. First, in paragraph 3.2.3 of Book 2 of its ethics guidelines, the MRC states in categorical terms that 'germline therapy should not be contemplated'. This is an unhappy choice of words, given that this ethics guideline would ostensibly render this present article – and the reader's act of reading it – unethical, merely because the content of this article *contemplates* germline therapy. Then, in paragraph 3.2.3.1, the MRC takes a less categorical position, by stating that 'gene modification of the human germline should not yet be attempted until such time that it is clearly sanctioned in South Africa'. This statement is comprehensively vague: Who must sanction human germline editing before it should be attempted, and when will such sanction be sufficiently clear? These positions stand in contrast with paragraph 2.17 of the same document, where the MRC takes a significantly more permissive approach to the genetic manipulation of embryos – which would include genome editing of embryos – when it states: 'Pre-embryo manipulation and research may yield valuable medical information … [T]he embryos should not be transferred to the uterus unless there is reasonable certainty that the manipulation carries no potential risks for the fetus'. The MRC is silent on risks to the prospective person (after birth) whose genome is to be edited.

We suggest that the antipathy towards human germline editing by the HPCSA and seemingly the MRC is problematic. There is no ostensible justification for placing an absolute prohibition on human germline editing. At best, considerations of safety and efficacy are good reasons for a temporary ban on clinical application, but not on research.[17] Also, safety and efficacy concerns may be of a transient nature, emphasising why any ban on clinical application should be temporary, and allow the opportunity for further research and clinical trials. Furthermore, the MRC's focus on the foetus – with no mention of the prospective person whose genome is to be edited, or his or her progeny – seems misplaced. Importantly, the MRC's requirement of 'no potential risks' is an unrealistic benchmark for when human reproduction is appropriate. Consider the 'potential risks' in normal conception: a recent study in the USA found that the frequency of foetuses potentially affected by a profound or severe genetic condition ranged from 94.5 to 392.2 per 100 000, depending on the ethnic group.[18] Accordingly, the standard of 'no potential risks' would necessitate the banning of all unprotected heterosexual sex.

Legislation

As a general rule, the relevant legislative instruments focus either on research, or on clinical application, allowing us to analyse these two regulatory spaces separately. We start with research, followed by clinical application.

Research on human germline editing

Ethics committee oversight of scientific research has become best practice globally. Section 73 of the *National Health Act* (NHA)[19] provides that every organisation that conducts 'health research' must have a health research ethics committee (HREC), and that this HREC must review research proposals and protocols, and grant approval where research proposals and protocols meet its ethical standards. Although not all human germline editing research would necessarily be health related, the NHA's broad definition of 'health research' is likely to include all human germline editing research within its ambit, and hence make HREC approval legally compulsory.

Given that research on human germline cells would require human research participants to donate germline cells, human germline editing research will trigger the application of the Regulations Relating to Research with Human Participants.[20] These Regulations provide that research involving human participants must: (a) comply with the Department of Health Ethics Guidelines at a minimum; (b) be responsive to health needs or priorities of the population, participating community or proposed participants; (c) have a valid scientific methodology and be likely to provide answers for the specific research questions posed; (d) include a favourable risk–benefit analysis; (e) ensure that the recruitment and selection process is just and fair; (f) be undertaken with appropriate consent processes; (g) undergo independent review by a registered HREC; (h) respect participants' rights, including but not limited to the rights to dignity, privacy, bodily integrity and equality; (i) make provision for compensation for research-related injury, for more than minimal risk research; and (j) be managed by a lead researcher, or person with similar standing or title, with suitable experience and qualifications. The Regulations also impose a wide array of further legal duties on researchers.

Research on human germline editing would also trigger the application of a second set of regulations made in terms of the NHA, namely the Regulations Relating to the Use of Human Biological Material.[21] These Regulations prescribe various legal requirements for 'genetic health research'. Given the broad definition of 'health research' in the NHA, and given that germline editing is undoubtably genetic in nature, research on human germline editing is likely to qualify as 'genetic health research'. The legal requirements for genetic health research prescribed by these Regulations include, inter alia, that such research must be done at an institution authorised as such in terms of the NHA, and that the institution must keep separate registers to record the genetic health research it conducts, and submit these registers to the Minister of Health by the end of March each year.

In addition, embryos would also be regulated by another provision of the NHA, namely section 57(4). This section provides that research on embryos within the first 14 days of embryonic development is permissible, subject to: (a) ministerial consent, (b) donor consent, and (c) an undertaking by the researcher to keep records of the research.

In the event that the research reaches a point, after properly designed preclinical studies,[13] where it is ready for human clinical trials – for an embryo with an edited genome to be transferred in utero for reproductive

purposes – section 71 of the NHA will apply. Although editing takes place on a germ cell prior to the prospective child's existence, the research does not stop with that act, but continues through the gestation and into the child's life. If the particular genome edit holds out the prospect of direct benefit to the child,[20] the research may only be conducted: (a) if it is in the best interests of the child; (b) in such manner and on such conditions as may be prescribed in regulations; and (c) with the consent of the parent or guardian of the child. Similar to a minor, the prospective child cannot consent. If the particular genome edit does not hold out the prospect of direct benefit to the child, but holds out the prospect of generalisable knowledge,[20] the consent of the Minister is required in addition to requirements (a) to (c) above. The NHA provides that the Minister may not give consent in circumstances where, inter alia, (i) the reasons for the consent to the research or experimentation by the parent or guardian are contrary to public policy; (ii) the research or experimentation poses a significant risk to the health of the minor; or (iii) there is some risk to the health or well-being of the minor and the potential benefit of the research does not significantly outweigh that risk.

Given our brief analysis above, we suggest that there is a robust formal legal regulatory environment for human germline editing research in South Africa. In addition, the Regulations Relating to Research with Human Participants also provide important substantive legal rules to determine whether to permit germline editing research, such as responsiveness to the health needs of the South African population, and a valid scientific methodology. However, a potential problem arises whenever there is deference to HRECs, as these bodies would be guided by the ethical rules contained in the leading South African ethics guidelines, which we suggest are problematic (as discussed above).

Clinical application of human germline editing

We now move to the clinical application of human germline editing, and first consider the *Medicines and Related Substances Control Act* (MRSCA).[22] If germline editing clinical applications fall within the regulatory ambit of the MRSCA, the South African Health Products Regulatory Authority (SAHPRA) can call on such applications to be registered. SAHPRA will review registration applications based on the criteria of safety, efficacy and quality – as determined by clinical trials. To fall within the regulatory ambit of the MRSCA, a germline editing application must qualify as either a 'medicine', or a 'medical device'. In brief, a medicine is a *substance* that is used for a medical purpose in humans, while a medical device is an *instrument* that is used for a medical purpose, but that does not achieve its primary intended action by pharmacological, immunological or metabolic means in the human body. Given that the court held that a bacterium qualifies as a *substance*,[23] it is not unrealistic that the biological components used in genome editing, such as guide RNA, enzymes and viral vectors, may qualify as substances. Similarly, given that a genome editing technology such as CRISPR-Cas9 and the viral vectors used to deliver it to target cells can be described as biological tools for precision work (and hence as instruments), it is not unrealistic that these biological tools may qualify as medical devices. However, it is uncertain whether germline editing applications would qualify as something that is used *in humans*, given that germline editing is performed on human germ cells, and not on human organisms – although germline editing will have an *effect* on human organisms. Lastly, while some potential clinical applications of human germline editing will have a medical purpose (which is broader than therapy, and includes prevention as well), other potential clinical applications of human germline editing will not have a medical purpose (such as increasing an individual's intelligence quotient or athleticism), and will therefore clearly fall outside the regulatory scope of the MRSCA. In sum, therefore: SAHPRA's legal mandate to regulate clinical applications of human germline editing (i) does not include non-medical applications, and (ii) is uncertain even in the case of applications with a medical purpose.

Let us now consider the NHA. A concept used in the NHA that can conceivably include the clinical application of human germline editing is 'reproductive cloning of a human being', which is defined in section 57(6)(a) of the NHA as:

> *'reproductive cloning of a human being' means the manipulation of genetic material in order to achieve the reproduction of a human being and includes nuclear transfer or embryo splitting for such purpose;*

The 'reproductive cloning of a human being' is declared illegal in section 57(1) of the NHA, and in section 57(5) is made a criminal offence punishable by a fine and up to 5 years' imprisonment.

Although the clinical application of human germline editing is not cloning as the term is generally understood, one can argue that because germline editing is a form of 'manipulation of genetic material', it falls within the NHA's definition of 'reproductive cloning of a human being'. If this argument is accepted, the clinical application of human germline editing would be illegal and a punishable criminal offence in South Africa.

On the other hand, applying the principles of statutory interpretation to section 57 leads to the conclusion that germline editing should *not* be included within the ambit of the definition of 'reproductive cloning of a human being'. Firstly, in South African law, our courts give effect to the apparent purpose of a provision rather than its plain, literal meaning[24,25] and there are a number of indicia that the purpose of section 57 is to specifically outlaw human reproductive cloning, including the heading of section 57 which reads 'Prohibition of reproductive cloning of human beings', and the repeated use of the word 'cloning' in section 57. Secondly, where a provision in a statute features the word 'include', the words after it define the general class of things that fall within the scope of the provision.[26] Thus, 'reproductive cloning of a human being' relates only to the general class of things that are defined by nuclear transfer and embryo splitting, namely cloning techniques. Accordingly, germline editing, which is not a cloning technique, would not be included in the definition of 'reproductive cloning of a human being'. Thirdly, where a provision in a statute is linked with a criminal sanction – as is the case with the definition of 'reproductive cloning of a human being' – the narrowest possible interpretation is preferred.[27] Applying this rule of our law, the definition of 'reproductive cloning of a human being' should only relate to cloning, and not more broadly to germline editing.

Although the application of the principles of statutory interpretation indicates that the legal position is that the clinical application of human germline editing is not prohibited, given the clumsy wording of the NHA, and the absence of case law on the subject, this position is all but certain.

Discussion

While there are some who hold that any form of genetic manipulation is morally reprehensible,[28] this is not a sentiment which is widely shared. Several prominent scientists and bioethicists have acknowledged that germline editing may be ethically permissible in at least some instances.[29] An overview of 61 ethics statements issued all over the world during the past 3 years shows that 54% of these statements take a position that germline editing should not be permitted, 30% are ambiguous or take no position, 11% are open to the possibility in the future, and 5% are open to further exploration.[30] However, in South Africa, the ethical and legal positions regarding human germline editing should first and foremost be informed by the values of the Constitution – most prominently dignity, equality and freedom.

The current South African ethics guidelines appear to be under-developed, and to be simply mimicking the position of most international ethics statements without due regard to South Africa's unique health-care needs, values, and existing legal regulatory environment. South Africa is dealing with epidemics, like TB and HIV, on a scale incomparable to the countries where some of the ethics statements referred to above originated. These diseases are undermining South Africans' quality of life, and hence their dignity; these diseases also disproportionally affect the poor, and as such exacerbate inequality. Moreover, South Africa is one of very few countries to explicitly protect the right to freedom of scientific research in its Constitution. In the present context, this right is in a mutually supporting relationship with the right to access to health care, the right to dignity and the right to equality. *Therefore, South Africa needs to be bolder (than other countries need to be) in seeking*

health-care solutions. Also consider that South Africa has a robust legal regulatory environment for health research, which would include research on human germline editing. This means that South Africa can be bold in seeking healthcare solutions *with confidence*.

When it comes to the legal regulation of the *clinical application* of human germline editing, the novelty of the technology creates uncertainty both in the context of the MRSCA and the NHA. Given the uniqueness of human germline editing (in contrast with, for instance, traditional medicines and medical devices) we suggest that *sui generis* legislation be developed to regulate the clinical application of human germline editing.

Recommended principles

In this section, we propose a set of guiding principles that we suggest should underlie and inform the regulation of human germline editing in South Africa. The first principle relates exclusively to research; the remaining four relate primarily to application, but can also be relevant to consider during the preceding research phase.

Principle 1: Human germline editing should be regulated, not banned

Given its potential to improve the lives of the people of South Africa, human germline editing should be regulated, not banned. Such research will qualify as health research, and be regulated by the relevant parts of the NHA and its regulations (discussed above). This would include a system of HREC oversight. When considering proposed research on human germline editing, an HREC should, we suggest, apply the same substantive criteria as with any other proposed health research that involves human participants who provide human biological material. One exception should be that, given the nature of germline editing (namely that the modifications made to the genome as part of germline editing are designed to be heritable and therefore may be passed on to future generations), consideration should be given to the potential long-term implications of the proposed research. This would be relevant not only with the risk–benefit analysis, but also with, inter alia, informed consent and, where relevant, community engagement.

Principle 2: Use the well-established standard of safety and efficacy

Following properly designed preclinical studies,[13] new human germline editing clinical applications should be subjected to clinical trials on humans – the same as with new medicines or medical devices. Clinical trial protocols would need to be designed mindful of the fact that germline editing is designed to be heritable. This means that the first human trials may have to monitor the trial participants over multiple years, perhaps even over generations. However, as with new medicines or medical devices, human germline editing clinical applications should only be made accessible to the public if they are proven to be safe and effective.

The safety and efficacy of germline editing will of course not affect any *existing* person; given the nature of germline editing, it will affect *prospective* persons. An important question in this context is whether one can consider the interests of prospective persons in South African law? The answer is yes: In *AB v Minister of Social Development* the Constitutional Court indeed considered the interests of *prospective* persons.[31] The Court based its eventual decision to uphold a legal prohibition on certain conduct on a factual finding that such conduct in the present would cause harm to prospective persons in the future.[31] (Note that the concept of the 'prospective person' is a mental construct of a person that may exist in future, and does not refer to the prenate.[32] Although an embryo or foetus may *become* the prospective person, an embryo or foetus cannot be *equated* to the prospective person.)[32] Accordingly, we suggest that, from a legal perspective, there is solid foundation to consider the safety and efficacy of germline editing for future generations.

Principle 3: Non-therapeutic uses of germline gene editing may be permissible

Non-therapeutic uses of germline editing are viewed by some as ethically problematic.[33,34] Such uses are often referred to as genetic 'enhancement' because they are viewed as nothing more than an endeavour to enhance people without valid moral justification. In Western bioethics in particular, 'enhancement' is viewed as morally reprehensible largely because it is seen as reminiscent of the state-sponsored eugenics programmes of early 20th-century Britain, America and Nazi Germany.[35] It is for this reason that ethics statements such as the one issued by the Association for Responsible Research and Innovation in Genome Editing (ARRIGE) in 2018, claim that genetic modification of the CCR5 gene to prevent children from contracting HIV is a genetic enhancement, and is therefore unnecessary and unethical.[36]

Whether this line of thought should have bearing on the South African position should be questioned, given that the equivalence between enhancement and eugenics is contested. As pointed out in the literature,[37,38] whereas state-enforced eugenic regimes used coersive means that *violated* procreative freedom in an attempt to create 'better people', individual uses of germline editing technologies, by contrast, *promote* procreative freedom by leaving their application up to individual choice. Furthermore, the assumption that there are no morally justifiable reasons for enhancement applications of germline editing is another issue worth questioning, given that several noteworthy bioethicists[29], and reputable ethics bodies such as the Nuffield Council on Bioethics, have opined that there may be circumstances where genetic enhancement would be ethically justifiable[11].

One such instance is the potential use of germline editing for the selection of desirable genetic traits in future offspring, in a way similar to choosing embryos using the information generated by pre-implantation genetic testing. It has been predicted that genetic selection against single-gene disorders (also known as Mendelian disorders) is one of the few likely candidates for which germline editing technology will be used, given that it provides parents who are both carriers for such diseases a means to have a child that is genetically related to them.[39] There is no reason, in principle, why germline editing could not be used for the selection of other single-gene traits – and perhaps even traits that are the product of the interaction of multiple genes. There is, further, no apparent reason why this would be deemed unacceptable in South Africa given that the genetic selection of gametes and embryos for non-medical reasons is permissible in our law (with the exception of sex selection, which may only take place to prevent a serious medical condition).[40] There is, in fact, a reason why this would be deemed to be a favourable alternative to genetic selection via pre-implantation genetic testing: Genetic selection using germline editing technology does not entail the destruction of multiple embryos that do not possess the desired genetic traits. This is an ethically compelling consideration for societies – such as South Africa – where the embryo is viewed by some as having a special moral status such that its destruction, in the context of medically assisted reproduction, ought to be avoided.[40]

While we do not make the claim, as some have, that this or any other application of germline editing is a basis for a moral duty to enhance future generations,[41] in our view a blanket prohibition on non-therapeutic applications of germline editing is inappropriate, as there may be ethically and legally defensible justifications for such applications. As such, we suggest that both therapeutic and non-therapeutic applications of germline editing may be permitted.

Principle 4: Respect parents' reproductive autonomy

Although human germline editing is an issue of broad societal interest, the choice to use germline editing – once it is safe and effective to use, and made available to the public – should be made by individual prospective parents. This is because, as recognised by the Nuffield Council on Bioethics,[11] the use of germline editing technology intersects with the high premium modern liberal democracies give to the need to respect the reproductive goals of persons seeking to become parents.[11] While some consider human germline editing an unprecedented

intrusion into the destiny of future generations, others have argued that it is in no way meaningfully different from other ways in which parents can influence their children.[42] Underlying these arguments is the claim that human germline editing falls within the ambit of socially accepted and legally protected interests of parents in making decisions relating to reproduction. It is for this reason that, in the American context, the National Academies of Sciences, Engineering, and Medicine 2017 report on human genome editing notes that: 'Access to heritable genome editing would be consistent with the broadest legal and cultural interpretations of parental autonomy rights in the United States.'[8] In South Africa, reproductive autonomy finds protection in section 12(2)(a) of the Constitution, which provides that: 'Everyone has the right to bodily and psychological integrity, which includes the right— (a) to make decisions concerning reproduction'. While this right has historically been interpreted primarily in the negative sense, i.e. in relation to the rights of individuals to choose *not to have children*, our courts have also acknowledged in recent years that reproductive autonomy also applies in the positive sense, i.e. in relation to rights of prospective parents to choose *to have a child* – including through the use of new reproductive technologies.[31] In so far as human germline editing technology may be viewed as a reproductive technology, reproductive autonomy extends to the use of germline editing in order to, for instance, allow parents who are both carriers of a genetic disorder to have a genetically related child who is free of that genetic disorder.

For these reasons, we suggest that in the event that clinical applications of germline editing become available to the public, prospective parents should be permitted to choose whether they wish to use such applications for their prospective children. The choices of prospective parents in this regard should not otherwise be restrained, unless it is a limitation which is reasonable and justifiable in an open and democratic society as per section 36 of the Constitution.

Principle 5: Promote the achievement of equality of access

A concern that is often raised in debates about human germline editing is that this new technology may possibly only be accessible to the rich, with the consequence of exacerbating existing inequalities in society – particularly in societies like South Africa given the wide gap between the rich and poor, and the lack of access to health care for the underprivileged.[43] One possible response is that although new technologies are often expensive initially, in time they typically become far less expensive; the early adopters of new technology pay a premium for it, and essentially fund the ongoing research and development of the technology to make it more accessible. From a legal perspective, the appropriate response to the concern about exacerbating inequality is the Constitutional Court dictum that measures to promote the achievement of equality call for 'equality of the vineyard not the graveyard'.[44] In other words, the solution to the concern about human germline editing exacerbating inequality cannot be to suppress the technology, as that would mean levelling down to the 'equality of the graveyard'; rather, if the state seeks to promote the achievement of equality in the context of human germline editing it must do so by levelling up to the 'equality of the vineyard'. As bioethicist John Harris has pointed out, the appropriate approach for a state that is genuinely concerned about novel technologies exacerbating inequality, would be to take measures to make these technologies as widely available as possible, and thereby remedying the inequality and promoting human flourishing.[37] To illustrate: universal health coverage of medically assisted reproduction is one strategy to promote access to new reproductive technologies[45], which has yielded positive results in some jurisdictions[46]. Such an approach may be viable in South Africa, which is currently in the process of implementing National Health Insurance.[47]

Conclusion

Human germline editing holds the promise of improving the lives of future generations. However, how can South African scientists work towards this aim in a milieu of regulatory uncertainty? Reform is needed. Instead of panicked reactions to He's germline editing actions, such as advocating moratoria, our proposed set of five guiding principles aims to provide a clear and realistic regulatory pathway for the South African science community to pursue human germline editing research, and eventually the clinical application thereof, in a responsible fashion aligned with the values of the South African Constitution.

For policymakers, we recommend that the relevant regulatory instruments (ethics guidelines and legislation) mentioned in this article be amended as indicated in Table 1. This represents the minimum action required to establish regulatory certainty and bring the South African regulatory environment in line with our proposed five guiding principles. A best-practice scenario would require, in addition, that the various South African ethics guidelines all incorporate the five principles.

Table 1: Recommended amendments to the current regulatory instruments for human germline editing

Regulatory instrument	Minimum required action
Health Professions Council of South Africa's ethics guidelines[15]	The bans on germline therapy research and on gene modification should be removed.
South African Medical Research Council's ethics guidelines[16]	(i) The ban on contemplating germline therapy should be removed. (ii) The requirement that genetically modified embryos may only be used for reproductive purposes if there is 'no potential risks for the fetus' should be replaced with the requirement that there is a favourable benefit–risk analysis for the prospective person, and for future generations.
South African Department of Health's ethics guidelines[14]	–
National Health Act (NHA)[19]	The definition of 'reproductive cloning of a human being' in Chapter 8, section 57(6)(a) should be amended by inserting the words 'genetically identical' before 'reproduction of a human being'. This will make it clear that reproductive cloning is banned, and not reproduction following any form of genetic manipulation.
Regulations Relating to Research with Human Participants[20]	–
Regulations Relating to the Use of Human Biological Material[21]	In the alternative to amending the ambivalent definition of 'reproductive cloning of a human being' in the NHA, the same objective would be accomplished by inserting a provision in the Regulations Relating to the Use of Human Biological Material to the effect that clinical applications of human germline editing should be permitted subject to health research ethics committee oversight of clinical trials, and regulation by the South African Health Products Regulatory Authority.
Medicines and Related Substances Control Act[22]	–
New legislation	New primary legislation should be developed – based on the five principles – to regulate the clinical application of human germline editing in South Africa.

For scientists who are intending to do human germline editing research, we recommend stalling plans until, at a minimum, the HPCSA and MRC ethics guidelines are amended as per Table 1. Amending these ethics guidelines will allow HRECs to consider and approve human germline editing research, and will open the door to ministerial approval of the use of embryos in terms of section 57(4) of the NHA. Note, however,

that such research would need to remain in vitro until the NHA or the Regulations Relating to the Use of Human Biological Material have been amended as per Table 1. Once such amendment(s) have been effected, HRECs can consider and approve clinical trials (i.e. using an embryo with a modified genome for reproductive purposes), and the Minister can permit such clinical trials involving children in terms of section 71 of the NHA.

Acknowledgements

We acknowledge the University of KwaZulu-Natal for funding through the African Health Research Flagship Grant and the South African National Research Foundation for funding (grant 116275).

Competing interests

We declare that there are no competing interests.

Authors' contributions

All authors contributed to the conceptualisation and writing of the article; D.T. was responsible for project leadership and funding acquisition.

References

1. Baltimore D, Berg P, Botchan M, Carroll D, Charo RA, Church G, et al. A prudent path forward for genomic engineering and germ line gene modification. Science. 2015;348(6230):36–38. https://doi.org/10.1126/science.aab1028
2. Normile D. CRISPR bombshell: Chinese researcher claims to have created gene-edited twins. ScienceMag. 2018 November 26 [cited 2019 Jul 18]. https://doi.org/10.1126/science.aaw1839
3. Association for Responsible Research and Innovation in Genome Editing. Common statement between the Association for Responsible Research and Innovation in Genome Editing (ARRIGE), the Genome Writers Guild (GWG) and the Japanese Society for Genome Editing (JSGE) [document on the Internet]. c2019 [cited 2019 Jul 18]. Available from: https://arrige.org/Common_statement_Arrige_GWG_JSGE.pdf
4. World Health Organization. WHO expert panel paves way for strong international governance on human genome editing [webpage on the Internet]. c2019 [cited 2019 Jul 18]. Available from: https://www.who.int/news-room/detail/19-03-2019-who-expert-panel-paves-way-for-strong-international-governance-on-human-genome-editing
5. Germline editing: time for discussion. Nat Med. 2015;21(4):295. https://doi.org/10.1038/nm.3845
6. Lander E, Baylis F, Zhang F, Charpentier E, Berg P. Adopt a moratorium on heritable genome editing. Nature. 2019;567:165–168. https://doi.org/10.1038/d41586-019-00726-5
7. Wolinetz CD, Collins FS. NIH supports call for moratorium on clinical uses of germline gene editing. Nature. 2019;567:175. https://doi.org/10.1038/d41586-019-00814-6
8. National Academies of Sciences, Engineering, and Medicine. Human genome editing: Science, ethics, and governance. Washington DC: The National Academies Press; 2017. https://doi.org/10.17226/24623
9. Ormond KE, Mortlock DP, Scholes DT, Bombard Y, Brody LC, Faucett WA, et al. Human germline genome editing. Am J Hum Genet. 2017;101(2):167–176. https://doi.org/10.1016/j.ajhg.2017.06.012
10. Nuffield Council on Bioethics. Genome editing – an ethical review [document on the Internet]. c2016 [cited 2019 Jul 18]. Available from: https://nuffieldbioethics.org/wp-content/uploads/Genome-editing-an-ethical-review.pdf
11. Nuffield Council on Bioethics. Genome editing and human reproduction: Social and ethical issues [webpage on the Internet]. c2018 [cited 2019 Jul 18]. Available from: https://nuffieldbioethics.org/publications/genome-editing-and-human-reproduction
12. Lagay F. Gene therapy or genetic enhancement: Does it make a difference? AMA J Ethics Virtual Mentor. 2001;3(2). https://doi.org/10.1001/virtualmentor.2001.3.2.gnth1-0102
13. Nordgren A. Designing preclinical studies in germline gene editing: Scientific and ethical aspects. J Bioeth Inq. 2019;16:559–570. https://doi.org/10.1007/s11673-019-09947-9
14. South African Department of Health (DoH). Ethics in health research: Principles, processes and structures. Pretoria: DoH; 2015.
15. Health Professions Council of South Africa. Guidelines for good practice in the health care professions. Booklet 14: General ethical guidelines for biotechnology research in South Africa [document on the Internet]. c2008 [cited 2020 Sep 10]. Available from: https://www.sada.co.za/media/documents/HPCSA_Booklet_14_Biotechnology_Research_in_SA.pdf
16. South African Medical Research Council. Guidelines on ethics in reproductive biology and genetic research [document on the Internet]. [updated 2002–2004; cited 2019 Jul 18]. Available from: http://www.mrc.ac.za/sites/default/files/attachments/2016-06-29/ethicsbook2.pdf
17. Macintosh KL. Heritable genome editing and the downsides of a global moratorium. CRISPR J. 2019;2(5):272–279. https://doi.org/10.1089/crispr.2019.0016
18. Haque IS, Lazarin GA, Kang HP, Evans EA, Goldberg JD, Wapner RJ. Modeled fetal risk of genetic diseases identified by expanded carrier screening. JAMA. 2016;316(7):734–742. https://doi.org/10.1001/jama.2016.11139
19. National Health Act 61 of 2003, South Africa.
20. The Republic of South Africa. Regulations Relating to Research with Human Participants GN R719 GG 38000 of 19 Sep 2014.
21. The Republic of South Africa. Regulations relating to the Use of Human Biological Material GN R177 GG 35099 of 2 Mar 2012.
22. Medicines and Related Substances Control Act 101 of 1965, South Africa.
23. Reitzer Pharmaceuticals (Pty) Ltd v Registrar of Medicines 1998 (4) SA 660 (T).
24. Bato Star Fishing (Pty) Ltd v Minister of Environmental Affairs and Tourism and Others 2004 (4) SA 490 (CC).
25. Natal Joint Municipal Pension Fund v Endumeni Municipality 2012 (4) SA 593 (SCA).
26. Minister of Safety and Security and Another v Xaba 2003 (2) SA 703 (D).
27. Rossouw v Sachs 1964 2 SA 551 (A).
28. Habermas S. The future of human nature. Cambridge: Polity; 2003.
29. Bosley KS, Botchan M, Bredenoord AL, Carroll D, Charo RA, Charpentier E, et al. CRISPR germline engineering – the community speaks. Nat Biotechnol. 2015;33:478–86. https://doi.org/10.1038/nbt.3227
30. Brokowski C. Do CRISPR germline ethics statements cut it? CRISPR J. 2018;1(2). https://doi.org/10.1089/crispr.2017.0024
31. AB v Minister of Social Development [2016] ZACC 43, 2017 (3) SA 570 (CC).
32. Jordaan DW. Stem cell research, morality, and law: An analysis of Brüstle v Greenpeace from a South African perspective. S Afr J Hum Rights. 2017;33(3):429–451. https://doi.org/10.1080/02587203.2017.1392839
33. The US National Academies of Sciences, Engineering and Medicine. International summit on human gene editing: A global discussion [webpage on the Internet]. c2015 [cited 2019 Nov 26]. Available from: https://www.nap.edu/catalog/21913/international-summit-on-human-gene-editing-a-global-discussion
34. US National Institute of Health. Statement on NIH funding of research using gene-editing technologies in human embryos [webpage on the Internet]. c2015 [cited 2019 Sep 24]. Available from: https://www.nih.gov/about-nih/who-we-are/nih-director/statements/statement-nih-funding-research-using-gene-editing-technologies-human-embryos
35. Comfort N. Can we cure genetic diseases without slipping into eugenics? The Nation. 2015 July 16 [cited 2019 Aug 24]. Available from: https://www.thenation.com/article/can-we-cure-genetic-diseases-without-slipping-into-eugenics/
36. Association for Responsible Research and Innovation in Genome Editing. Statement from ARRIGE steering committee on the possible first gene-edited babies [document on the Internet]. c2018 [cited 2019 Sep 01]. Available from: http://arrige.org/ARRIGE_statement_geneeditedbabies.pdf
37. Harris J. Enhancing evolution: The ethical case for making better people. Princeton, NJ: Princeton University Press; 2010
38. Bailey R. Liberation biology: The scientific and moral case for the biotech revolution. Amherst, NY: Prometheus; 2005.
39. Greely HT. Human germline genome editing: An assessment. CRISPR J. 2019;2:253–265. https://doi.org/10.1089/crispr.2019.0038
40. The Republic of South Africa. Regulations Relating to Artificial Fertilisation of Persons GN R175 of 2012.
41. Savulescu J, Pugh J, Douglas T, Gyngell C. The moral imperative to continue gene editing research on human embryos. Protein Cell. 2015;6(7):476–479. https://doi.org/10.1007/s13238-015-0184-y

42. Silver LM. Remaking Eden: How genetic engineering and cloning will transform the American family. New York/London: Ecco; 2007 .

43. Kotzè M. Human genetic engineering in the South African context with its inequalities: A discourse on human rights and human dignity. Scrip. 2014;113(1)1–11. https://doi.org/10.7833/113-0-722

44. Minister of Home Affairs v Fourie; Lesbian and Gay Equality Project v Minister of Home Affairs [2005] ZACC 19, 2006 (1) SA 524.

45. Hemel D. The case for universal health coverage for in vitro fertilization. The Seattle Times. 2018 November 15 [cited 2020 Jan 02]. Available from: https://www.seattletimes.com/opinion/the-case-for-universal-health-coverage-for-in-vitro-fertilization

46. Vélez MP, Connolly MP, Kadoch IJ, Phillips S, Bissonnette S. Universal coverage of IVF pays off. Hum Reprod. 2014;29(6)1313–1319. https://doi.org/10.1093/humrep/deu067

47. National Health Insurance Bill. B11-2019, South Africa.

AUTHORS:
Sisa Pazi[1]
Chantelle M. Clohessy[1]
Gary D. Sharp[1]

AFFILIATION:
[1]Department of Statistics, Nelson Mandela University, Port Elizabeth, South Africa

CORRESPONDENCE TO:
Gary Sharp

EMAIL:
Gary.sharp@mandela.ac.za

DATES:
Received: 20 Apr. 2020
Revised: 18 May 2020
Accepted: 13 June 2020
Published: 29 Sep. 2020

HOW TO CITE:
Pazi S, Clohessy CM, Sharp GD. A framework to select a classification algorithm in electricity fraud detection. S Afr J Sci. 2020;116(9/10), Art. #8189, 7 pages. https://doi.org/10.17159/sajs.2020/8189

ARTICLE INCLUDES:
☒ Peer review
☐ Supplementary material

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

EDITOR:
Amanda Weltman

KEYWORDS:
electricity fraud detection, confusion matrix, classification algorithms

FUNDING:
SASA/NRF Crisis in Academic Statistics (grant no. 94756); Nelson Mandela University.

A framework to select a classification algorithm in electricity fraud detection

In the electrical domain, a non-technical loss often refers to energy used but not paid for by a consumer. The identification and detection of this loss is important as the financial loss by the electricity supplier has a negative impact on revenue. Several statistical and machine learning classification algorithms have been developed to identify customers who use energy without paying. These algorithms are generally assessed and compared using results from a confusion matrix. We propose that the data for the performance metrics from the confusion matrix be resampled to improve the comparison methods of the algorithms. We use the results from three classification algorithms, namely a support vector machine, *k*-nearest neighbour and naïve Bayes procedure, to demonstrate how the methodology identifies the best classifier. The case study is of electrical consumption data for a large municipality in South Africa.

Significance:

- The methodology provides data analysts with a procedure for analysing electricity consumption in an attempt to identify abnormal usage.
- The resampling procedure provides a method for assessing performance measures in fraud detection systems.
- The results show that no single metric is best, and that the selected metric is dependent on the objective of the analysis.

Introduction

Revenue that is lost due to the difference between electricity supplied and electricity purchased is partitioned into two classes. The first class resulting from transmission and other infrastructural limitations is labelled as technical losses, whilst the second class, the majority of which are a result of meter tampering or bypassing, is labelled as non-technical losses[1-3] (NTL). Estimates of losses worldwide are in the billions[4,5] of US dollars and suppliers of electricity have expressed concern over these losses and the sustainability of the supply[6,7].

The literature related to fraudulent electricity losses is detailed with the first traceable case as early as the 19th century claiming, 'unprincipled persons had availed themselves of the opportunity to steal electricity'[8,9]. More recent literature is a result of the computational hardware and software developments over the last two decades. Galvan et al.[10] brought to the fore statistical methods for identifying 'abnormal' consumer behaviour in the electrical domain. This has seen fraud detection systems from finance, banking and insurance being applied to the electricity domain. The review by Messinis and Hatziargyriou[11] of the methods applied to detect electricity theft provides compelling evidence that the research domain is exciting and extensive. The review informs researchers of the types of data used in fraud detection, the algorithms that have been proposed and clarifies performance metrics for comparing the algorithms.

A comprehensive list of the classifiers used in electricity fraud detection systems can be found in Messinis and Hatziargyriou[11]. Notably these include support vector machines (SVM), naïve Bayesian (NB) methods and *k*-nearest neighbour (*k*-NN) classifiers – the algorithms used in this study. The field has not stagnated; recent computational methods include convolutional neural networks[3] and ensemble-based classifiers[7] whilst time series methods have been explored[2]. In many of these studies, the common methodological approach is to assess the classifier by considering results from a test data set captured in a confusion matrix summary and reported as a performance measure. Several performance measures are used in the literature and for the most part are defined to reflect accuracy and precision of the classifiers.

The confusion matrix is a 2 x 2 table summarising the predicted versus the actual frequency counts for a binary classification model. The table, used in the financial sector to identify fraudulent customers, is best suited to data sets for which the number of observations in the two classes are similar or moderately similar. Ideally for a model, the predicted results match the actual counts. In the confusion matrix, this implies that the frequencies of the true positives are the same as the actual positives and the frequencies of the true negatives are the same as the actual negatives. In addition, the false negatives and false positives should be zero. The confusion matrix, in Table 1, was used in studies by Messinis et al.[12] and Li et al.[3] to assess classification models, whilst Guo et al.[13] addressed the problem for data sets which are highly imbalanced.

Our research study had a dual objective. The first objective was to use a resampling approach to assess the performance of the classifier using a data set from the Nelson Mandela Bay Municipality. The second objective was to add to the South African literature on fraud detection in the electrical domain.

Table 1: A 2 × 2 confusion matrix for two classes

True class	Predicted class		Total
	Positive	Negative	
Positive	True positive (TP)	False negative (FN)	Actual positive
Negative	False positive (FP)	True negative (TN)	Actual negative
Total	Predicted positive	Predicted negative	**Total counts**

Literature review

Electricity fraud detection methods were adapted to a large extent from the methods used to detect fraudulent activity in the banking, insurance and telecommunications sector. Bolton and Hand[14] review classification models used for binary identification, albeit in the finance sector. Although the type of data and the ratio of fraudulent activity differ between the financial and electrical domains, the methods are similar, hence they are adapted accordingly.

The last two decades have seen considerable research directed towards electricity fraud. From the earlier work by Galvan et al.[10] who evaluated electricity usage in the Spanish farming sector, to Davidson[15] and Fourie and Calmeyer[16] who introduced NTL research into the South African context, to the behavioural identification of Hu et al.[17], the opportunity for research is extensive. Examples of research using fraud detection classification in electricity include Nizar et al.[18] who used a NB classifier and decision tree algorithm to assess the consumption load profile of customers at different time intervals. Nagi et al.[19] used a SVM classifier to detect fraud for a power system in Malaysia whilst in India, Depuru et al.[20] used smart meter data in their SVM classification study. Coma-Puig et al.[21] used *k*-NN and SVM classifiers, amongst others, to evaluate electricity data from Spain, whilst Li et al.[3] used several classifiers, including a hybrid random forest classifier, to evaluate data from Ireland. In each study, performance metrics were used to assess which classifier could be considered the best. For a review on classification algorithms in the electricity fraud detection sector the reader is referred to the study by Messinis and Hatziargyriou[11].

To assess classifiers, predicted data are summarised as a confusion matrix and often reported as a performance measure. Messinis and Hatziargyriou[11] clarify the pitfalls of using a single performance measure to assess a classifier. As an example, accuracy is claimed to be the most commonly used metric from a confusion matrix. Accuracy is defined as the percentage of correct classifications in total. However, this statistic summarises both classes simultaneously and if the data set is imbalanced, the correct classification of the larger class could distort the results. As a result, researchers have chosen to include several performance measures when reporting the assessment of a classifier. Coma-Puig et al.[21] used the performance measures recall and *f*-measure whilst Ghori et al.[7] and Li et al.[3] opted to include precision as one of their measures. There is little consensus on which metric is best; however, there is consensus that more than one metric should be reported. Messinis and Hatziargyriou[11] list seven performance metrics used in their review. Their tabulated summary shows that accuracy is the most common metric, followed by detection rate, precision and false positive rate. Definitions and justifications of metrics used in this study are defined in the methodology.

Eskom, a state-owned utility, is the primary energy supplier in South Africa, and generates almost 95% of South Africa's energy.[22] Eskom initiated a campaign called Operation Khanyisa in an effort to combat electricity theft in South Africa.[6] The campaign reported that Eskom lost approximately USD300 million (ZAR5 billion) in 2016 as a consequence of NTL. Losses to this extent are unsustainable and place at risk the service provider's ability to ensure a steady supply of power.

Apart from Davidson[15] and Fourie and Calmeyer[16], there has been limited research focused on identifying electricity fraud in the South African sector. Doorduin et al.[23] used a simulation experiment to demonstrate how to identify customers whose consumption was irregular. Other industries within the South African context have developed classification models specific to their needs. Examples include Stalmans and Irwin[24] who used classification algorithms in cyber-computing to identify malware infections on a local electricity network. Breed and Verster[25] focused on the banking sector whilst Govender et al.[26] applied the methods in the forecasting of solar irradiance. The scarcity of research in the South African electricity fraud detection sector provides the opportunity for this study.

The data

The data set was obtained from a company that was contracted by the Nelson Mandela Bay Municipality to identify possible fraudulent activity in an effort to reduce NTL in the municipality. The data consisted of customers' historical electricity consumptions in kilowatt-hours (kWh) for 24 months from March 2013 to February 2015. Each customer's consumption pattern was categorised as either honest or fraudulent, based on an inspection undertaken by onsite inspectors. Fraudulent customers were defined as those customers where evidence of meter tampering was found by the on-site inspection whilst the remainder were defined as honest. In terms of the confusion matrix, positive values identify the fraudulent consumers and negative values identify the honest consumers.

The data set consisted of 3156 customers, of which 2420 (77%) were categorised as honest and 736 (23%) as fraudulent. A sample of this data set is shown in Table 2. The data were pre-processed and sorted for analytical routines using freeware R v.3.6.1[27] and licensed TIBCO software Statistica[28] v.13.

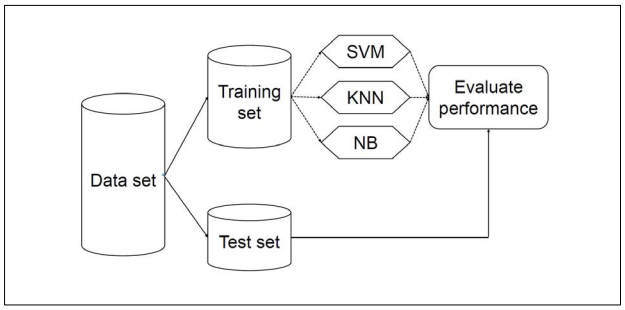
Table 2: A sample of the data set

Customer	Mar-13 (kWh)	Apr-13 (kWh)	Feb-15 (kWh)	Class label
1	222.3	209	148.4	Fraudulent
2	240.5	192.5	334.4	Honest
3	459.3	419.1	711	Honest
4	128.5	128.5	121.5	Fraudulent
5	957.2	889.2	150.7	Honest
3156	134.9	158	83.6	Honest

Although not all NTLs are a result of dishonest activities, for the purposes of this study, NTLs are collectively referred to as fraudulent activities. This definition allows for the use of a data-oriented supervised approach as the data available for analytics were the monthly electrical usage per customer as well as whether the consumer was labelled as an honest client or a fraudulent client.

Methodology

Data-oriented supervised methods require that the source data can be partitioned into two classes as defined by the research study. In electricity fraud, the data must be partitioned into either fraudulent or non-fraudulent (honest) classes whilst the independent terms can include several factors. There are two methods available to collect consumer electricity usage: older meters or those referred to as smart meters. Smart meters have the capability to record electrical use at different time intervals (i.e. hourly, daily, monthly) as well as other information such as the location or area of the meter and the billing costs of the user.[20] Older meters have less reliable data in that consumption readings are often only available on a monthly basis and, in most cases, are obtained by manual inspection, monthly estimation or consumer feedback. A limitation to this study is that the data available are predominantly from older meter readings and are subject to human capturing errors. In this study, three classification algorithms were assessed on the case study data following the procedure in Figure 1.



SVM, support vector machine; KNN, k-nearest neighbour classifier; NB, naïve Bayesian

Figure 1: General framework of the methodology

The data were partitioned into a training set and a test set, after which the training data were used by classifiers SVM, *k*-NN and NB to determine estimates for the prediction model. The model was then assessed by using the test data set to predict the classification of a consumer based on their electricity consumption and then the results were compared against the actual classification. These data were then summarised in a confusion matrix. The sequential procedure followed was:

- Step 1: Pre-process the data by cleaning and sorting as required. Once the data are in the format required by the software for analytical requirements, proceed to Step 2.
- Step 2: Randomly partition two thirds of the sample to a training data set and the remaining one third to a test data set.
- Step 3: Estimate the parameters of the model using the training data.
- Step 4: Test the performance of a fitted model on the test data set as follows:
  - Predict the class membership for everyone in the test set.
  - Compute a confusion matrix.
  - Calculate performance metrics from results in the confusion matrix.
  - Repeat Steps 2 to 4 p times to obtain p confusion matrices and p estimates for each metric.
- Step 5: Summarise the performance metrics and assess the statistics inferentially.

Four performance measures – accuracy, detection rate, precision and true negative rate – were used in this research. Accuracy, detection rate and precision are used extensively in the literature, and therefore it was important that they be included in this study. True negative rate is selected as the fourth measure as it complements precision by providing information to the researcher about the classifier's ability to correctly select the second class of consumers relative to the actual numbers in that class. The measures are defined using the notation from Table 1.

$$\text{Accuracy} = \frac{(TP+TN)}{(\text{Total counts})},$$

$$\text{Detection rate (Sensitivity)} = \frac{TP}{(\text{Actual positives})},$$

$$\text{Precision} = \frac{TP}{(\text{Predicted positives})}, \text{and}$$

$$\text{True negative rate (Specificity)} = \frac{TN}{(\text{Actual negatives})}.$$

These measures are summarised for the p number of iterations and thereafter assessed inferentially. We used analysis of variance (ANOVA) and Bonferroni post-hoc comparisons to assess the performance metrics. The ANOVA methodology compares the three mean responses for the classifiers for the four performance measures. In cases where the three mean responses were found to differ, the post-hoc comparisons were used to identify where the differences were. The final step in the methodology was the use of Cohen's D to assess the practical significance of the post-hoc comparisons.

Results and discussion

Each classifier was trained on the same folds of training data and thereafter the corresponding test data sets were used to obtain results for the confusion matrix. Each randomly selected training set consisted of 2104 consumers, of which approximately 77% were categorised as honest and 23% as fraudulent. The optimal parameters for the SVM classifier were $\gamma = 1.3542$ and $C = 2.1639$, while for the *k*-NN classifier, *k* was 20 and the optimal distance metric was the City Block.

A total of $p = 500$ iterations was used to obtain 500 randomly obtained test set folds, with each fold containing 1052 consumers, of which approximately 77% were categorised as honest and 23% were fraudulent. For the 500 test data sets, 500 confusion matrices were calculated for each classifier. The first iteration of the SVM classifier test data set is shown in Table 3. The results show that, of the 264 fraudulent customers, the classifier was able to correctly predict 196 customers – a precision of 74.4%. Similarly, the accuracy of the classifier was 88.2%.

Table 3: Predicted classes for the first iteration of the support vector machine classifier

True class	Predicted class		Total
	Positive (fraudulent)	Negative (honest)	
Positive (fraudulent)	196	68	264
Negative (honest)	26	762	788
Total	222	830	**1052**

Using the results from each iteration for each algorithm, 500 confusion matrices were obtained, and the four performance measures were summarised. The summary statistics for each metric and each classifier are shown in Table 4.

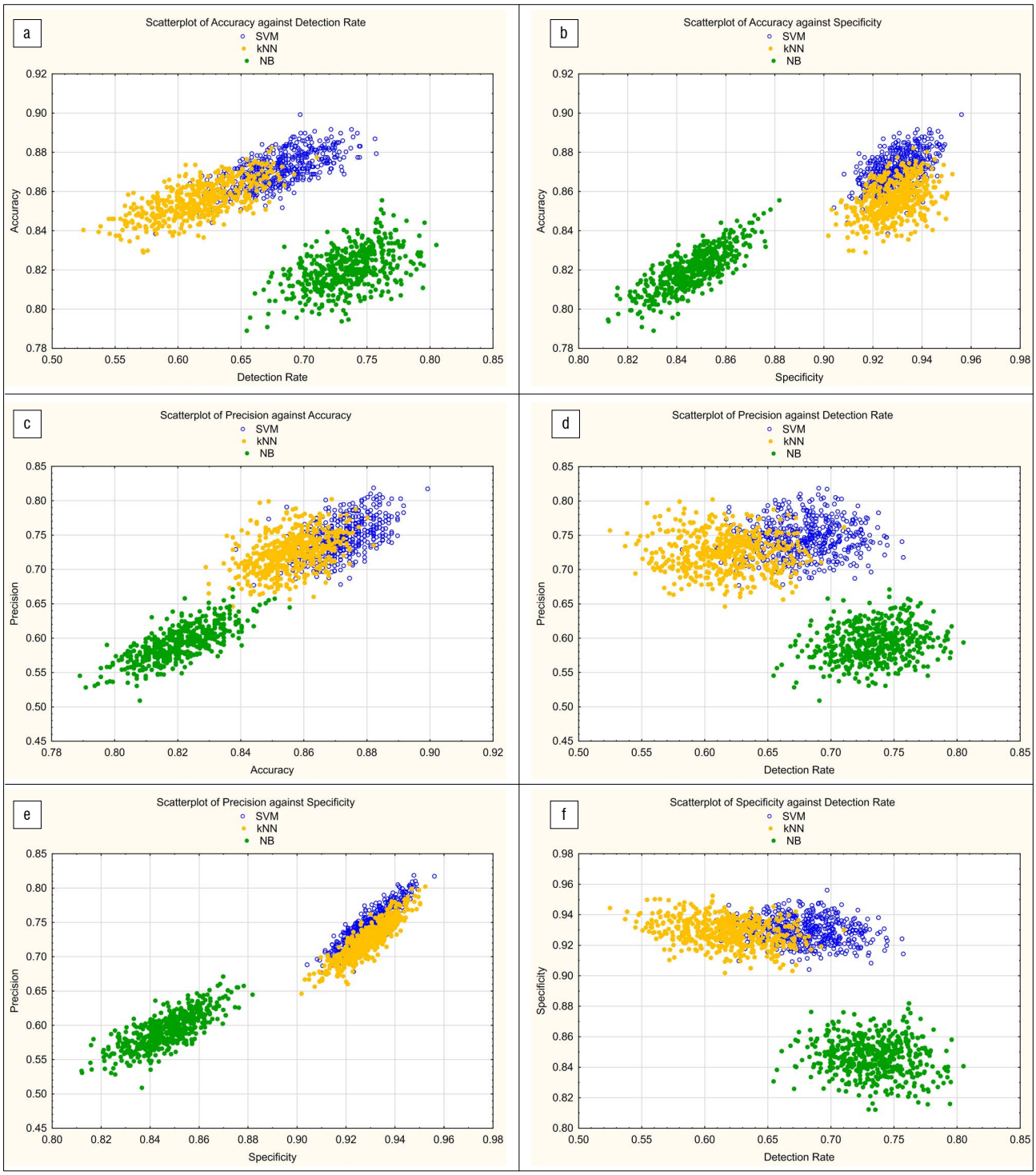
Table 4: Performance measures for the three classifiers

Metric	Classifier	Mean	s.d.	Minimum	Maximum
Accuracy	SVM	87.12%	0.88%	83.8%	89.9%
	k-NN	85.58%	0.91%	82.9%	88.2%
	NB	82.02%	1.04%	78.9%	85.6%
Detection rate	SVM	67.98%	2.77%	58.2%	75.7%
	k-NN	61.68%	3.09%	52.5%	71.0%
	NB	73.38%	2.64%	65.5%	80.5%
Precision	SVM	74.56%	2.65%	67.7%	81.9%
	k-NN	72.43%	2.72%	64.6%	80.2%
	NB	59.23%	2.52%	50.9%	67.1%
True negative rate	SVM	92.95%	0.84%	90.4%	95.6%
	k-NN	92.86%	0.88%	90.2%	95.2%
	NB	84.65%	1.19%	81.2%	88.2%

SVM, support vector machine; k-NN, k-nearest neighbour; NB, naïve Bayesian

For each performance measure, the higher the value (%), the better the analytical ability of the classifier to correctly determine the metric. The results in Table 4 indicate that the NB method is the least successful classifier of accuracy, precision and true negative rate, but is the best metric for detection rates. The metric mean is the lowest for the three metrics – accuracy, precision and true negative rate – whilst the variability estimate (the standard deviation) is the largest for two of the three metrics.

This observation is more apparent in three of the graphical illustrations in Figure 2. In Figure 2b, 2c and 2e, the clustering of the estimates from each iteration demonstrate that the NB algorithm performs poorly in comparison to the *k*-NN and SVM classifiers. A different pattern emerges for comparisons with the metric detection rate, where NB outperforms both SVM and *k*-NN. These graphical summaries indicate that no single classifier is the best for all four metrics; the evidence indicate that the classifiers have different abilities. In addition, the plots indicate that there is some overlap of results for the classifiers SVM and *k*-NN, implying that using both classifiers does not have much benefit. Arguably for data analytic purposes, only one of these two classifiers needs to be used in this domain.



SVM, support vector machine; kNN, k-nearest neighbour classifier; NB, naïve Bayesian

Figure 2: Scatter plots of classifier comparisons in terms of performance metrics.

Table 5 summarises the ANOVA results for each metric. For all four metrics, the *p*-value is very small (<0.01), providing inferential evidence supporting the claim that each performance measure's mean responses for the three classifiers are not equal. The Bonferroni post-hoc comparisons of the differences in performance measures mean responses between two classifiers is also provided in Table 5. Based on the findings with *p*-values <0.01, there is sufficient evidence to believe that the mean responses per metric for the classifiers differ. The exception is the true negative rate mean responses for the classifiers SVM and *k*-NN ($p=0.43$). These findings indicate that at least one classifier is statistically superior to the other(s).

Table 5: ANOVA and post-hoc summary results for each performance measure

ANOVA summary			Bonferroni *p*-values of differences between means		
Metric	***F*-statistic**	***p*-value**	**SVM vs *k*-NN**	**SVM vs NB**	***k*-NN vs NB**
Accuracy	3836.7	0.000	0.000	0.000	0.000
Detection rate	2129.6	0.000	0.000	0.000	0.000
Precision	4971.3	0.000	0.000	0.000	0.000
True negative rate	11692.2	0.000	0.434	0.000	0.000

SVM, support vector machine; k-NN, k*-nearest neighbour; NB, naïve Bayesian*

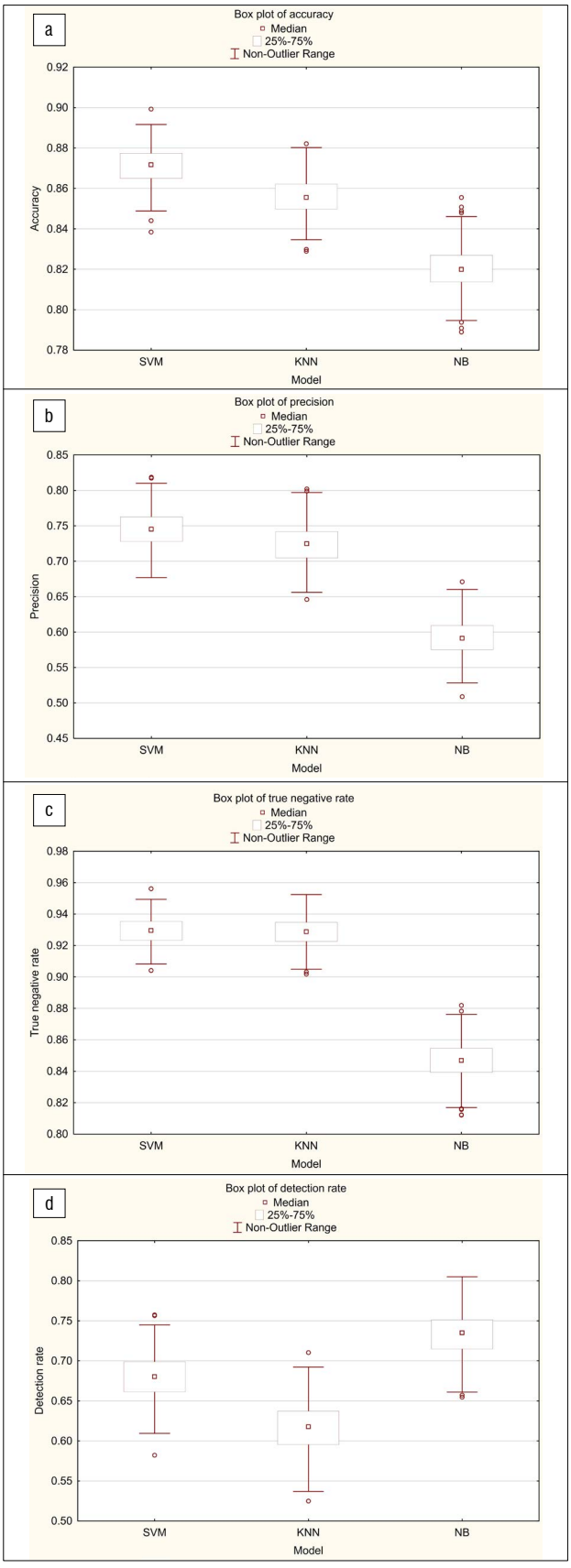
The results in Table 6 report the practical significance of these post-hoc tests using Cohen's D. Of the 12 comparisons, 10 report Cohen's D values exceeding 0.80, while another has a Cohen's D of 0.79. These results are referred to as a large effect and lend evidence to the claim that they are practically significant. The sole small effect is observed for the true negative rate mean responses for the classifiers SVM and *k*-NN, implying that this response is both practically and statistically insignificant.

Table 6: Practical significance for differences between means

Mean differences	Metric	Cohen's D	Rule of thumb
SVM vs *k*-NN	Detection rate	2.150	Large effect
	True negative rate	0.105	Small effect
	Accuracy	1.727	Large effect
	Precision	0.794	Medium effect
SVM vs NB	Detection rate	1.996	Large effect
	True negative rate	8.038	Large effect
	Accuracy	5.306	Large effect
	Precision	5.921	Large effect
***k*-NN vs NB**	Detection rate	4.074	Large effect
	True negative rate	7.824	Large effect
	Accuracy	3.646	Large effect
	Precision	5.028	Large effect

SVM, Zupport vector machine; k-NN, k*-nearest neighbour; NB, naïve Bayesian*

A graphical approach was used to identify which of the classifiers performed the best collectively for the four performance measures. The box plots in Figure 3a–d illustrate the four performance metrics for each classifier in relation to each other. SVM outperforms *k*-NN for all four metrics, indicating it is the better classifier of the two in this study. The box plots in Figure 3a–c which illustrate the metrics accuracy, precision and true negative rate, respectively, show that the SVM classifier is considerably better than the NB classifier, whilst in Figure 3d the opposite is observed for the detection rate metric. Not only is NB better than SVM, it outperforms *k*-NN by some margin. These observations lend support to the claim that, collectively, the SVM classifier is the better model for this data set, whilst the other two classifiers give conflicting results.



SVM, support vector machine; KNN, k*-nearest neighbour classifier; NB, naïve Bayesian*

Figure 3: Box plots of classifier comparisons in terms of performance metrics

In summary, the three classifiers were compared using four performance measures: accuracy, precision, detection rate and true negative rate. The results for these measures differ significantly as shown graphically and inferred using ANOVA. As the results indicate, no single classifier outperforms all others for every metric. Data analysts need to decide for themselves which performance measure is more relevant to their objectives. As an example, if detection rate is crucial to the study, then it would be useful to include the NB classifier. Alternatively, if an analyst needed a single classifier for these four performance measures, the SVM algorithm is recommended.

Conclusion

This research introduces statistical learning techniques as a method to identify electricity fraud in South Africa. The methodology involved the use of the three classification modelling approaches which were assessed using a test data set. The algorithms were compared using four performance measures which have been applied to fraud detection studies in finance and electricity theft. This methodology contributes to the literature by demonstrating a resampling approach to compare the performance measures. It is important to realise that there is considerable variability when partitioning data into training and test sets and once-off results can be misleading. Adopting a simple resampling approach can provide more clarity on the ability of the classifier to detect electricity fraud. In South Africa, municipalities can improve their revenue stream by identifying fraudulent clients and ensuring that the lost revenue is recovered. This in turn will allow the municipalities to reimburse Eskom for the electricity consumed. Adopting this methodology for fraud detection will allow municipalities to target potential fraudulent customers and reduce the number of manpower hours that would otherwise be required to conduct random on-site inspections.

Acknowledgements

This work is based on data supplied by Amat Security and Cleaning Services (Pty) Ltd on behalf of the Nelson Mandela Bay Municipality. Financial support was provided in part by the SASA/NRF Crisis in Academic Statistics (grant no. 94756) and the Nelson Mandela University Research Department.

Competing interests

We declare that there are no competing interests.

Authors' contributions

S.P. undertook the research as a master's student and conducted the data cleaning and statistical analysis, and wrote the first draft of the paper. G.D.S. and C.M.C. were responsible for supervision and revising the original paper. All authors were involved in the conceptualisation and methodology, and gave input into multiple revisions of the paper.

References

1. Monedero IL, Biscarri F, Leon C, Biscarri JY, Millan R. MIDAS: Detection of non-technical losses in electrical consumption using neural networks and statistical techniques. In: Proceedings of the International Conference on Computational Science and its Applications. Berlin: Springer; 2006. p. 725–734. https://doi.org/10.1007/11751649_80
2. Villar-Rodriguez E, Del Ser J, Oregi I, Bilbao MN, Gil-Lopez S. Detection of non-technical losses in smart meter data based on load curve profiling and time series analysis. Energy. 2017;137:118–128. https://doi.org/10.1016/j.energy.2017.07.008
3. Li S, Han Y, Yao X, Yingchen S, Wang J, Zhao Q. Electricity theft detection in power grids with deep learning and random forests. J Electr Comput Eng. 2019; Art. #4136874, 12 pages. https://doi.org/10.1155/2019/4136874
4. PowerNews. Municipalities, utilities meet in bid to protect dwindling revenues [webpage on the Internet]. c2013 [cited 2015 Mar 13]. Available from: http://www.powernews.co.za/news/entry/municipalities-utilities-meet-in-bid-to-protect-dwindling-revenues
5. T & D World Magazine. Home > Smart Grid > India to spend $21.6 billion on smart grid infrastructure by 2025 [webpage on the Internet]. c2015 [cited 2016 Jun 30]. Available from: http://tdworld.com/smart-grid/india-spend-216-billion-smart-grid-infrastructure-2025
6. Operation Khanyisa Partners [homepage on the Internet]. c2017 [cited 2017 Jun 21]. Available from: www.operationkhanyisa.co.za
7. Ghori KM, Abbasi RA, Awais M, Imran M, Ullah A, Szathmary L. Performance analysis of different types of machine learning classifiers for non-technical loss detection. IEEE Access. 2019;8:16033–16048. https://doi.org/10.1109/ACCESS.2019.2962510
8. Daily Yellowstone Journal, 1886. People who steal Edison's electricity. In: Glauner P, Glaeser C, Dahringer N, Valtchev P, State R, Duarte D. ResearchGate 2017; 9 pages. Available from: https://www.researchgate.net/publication/325297875
9. Glauner P, Meira JA, Valtchev P, State R, Bettinger F. The challenge of non-technical loss detection using artificial intelligence: A survey. Int J Comput Int Sys. 2017;10(1):760–775. https://doi.org/10.2991/ijcis.2017.10.1.51
10. Galvan JR, Elices A, Muniz A, Czernichow, Sanz-Bobi MA. System for detection of abnormalities and fraud in consumer consumption. Paper presented at: The 12th Conference on the Electric Power Supply Industry; 1998 November 2–6; Pattaya, Thailand.
11. Messinis GM, Hatziargyriou ND. Review of non-technical loss detection methods. Electric Power Systems Res. 2018;158:250–266. https://doi.org/10.1016/j.epsr.2018.01.005
12. Messinis G, Dimeas A, Rogkakos V, Andreadis K, Menegatos I, Hatziargyriou N. Utilizing smart meter data for electricity fraud detection. Paper presented at: First South East European Regional CIGRE Conference; 2016 June 7–8; Portoroz, Slovenia. p. 4–21.
13. Guo X, Yin Y, Dong C, Yang G, Zhou G. On the class imbalance problem. IEEE Fourth Int Conf Nat Comput. 2008:192–201. https://doi.org/10.1109/ICNC.2008.871
14. Bolton RJ, Hand DJ. Statistical fraud detection: A review. J Stat Sci. 2002;17:235–255.
15. Davidson IE. Evaluation and effective management of non-technical losses in electrical power networks. IEEE Africon, Pretoria, South Africa. 2002;473–477.
16. Fourie JW, Calmeyer JE. A statistical method to minimize electrical energy losses in a local electricity distribution network. In: Proceedings of the IEEE Africon; 2004 September 15–17; Gaborone, Botswana. IEEE; 2004. p. 667–673.
17. Hu W, Yang Y, Wang J, Huang X, Cheng Z. Understanding electricity-theft behavior via multi-source data. Paper presented at: WWW '20; 2020 April 20–24; Taipei, China. https://doi.org/10.1145/3366423.3380291
18. Nizar AH, Dong ZY, Zhao JH. Zhang P. A data mining based NTL analysis method. In: Proceedings of the IEEE Power Engineering Society General Meeting; 2007 June 24–28; Tampa, Florida, USA. IEEE; 2007. p. 1–8. https://doi.org/10.1109/PES.2007.385883
19. Nagi J, Mohammad AM, Yap KS, Tiong SK, Ahmed SK. Non-technical loss analysis for detection of electrcity theft using support vector machines. In: Proceedings of the IEEE 2nd International Conference on Power and Energy; 2008 December 1–3; Johor Bahru, Malaysia. IEEE; 2008. p. 907–912. https://doi.org/10.1109/PECON.2008.4762604
20. Depuru SSSR, Wang L, Devabhaktuni V. Support vector machine based data classification for detection of electricity theft. IEEE/PES Power Systems Conf Expos. 2011:1–8
21. Coma-Puig B, Carmona J, Gavalda R, Alcoverro S, Martin V. Fraud detection in energy consumption: A supervised approach. IEEE Int Conf Data Sci Adv Anal. 2016. Available from: https://www.cs.upc.edu/~gavalda/papers/dsaa2016.pdf
22. Pazi S. Statistical methods for the detection of non-technical losses: A case study for the Nelson Mandela Bay Muncipality [unpublished MSc dissertation]. Port Elizabeth: Nelson Mandela University; 2016.
23. Doorduin WA, Mouton HT, Herman R, Beukes HJ. Feasibility study of electricity theft detection using mobile remote check meters. In: Proceedings of the 7th IEEE Africon; 2004 September 15–17; Gaborone, Botswana. IEEE; 2004. p. 373–376. https://doi.org/10.1109/AFRICON.2004.1406697

24. Stalmans E, Irwin BA. A framework for DNS based detection and mitigation of malware infections on a network. In: Proceedings of the IEEE Information Security for South Africa; 2011 August 15–17; Johannesburg, South Africa. IEEE; 2011. p. 1–8. https://doi.org/10.1109/ISSA.2011.6027531

25. Breed DG, Verster T. The benefits of segmentation: Evidence from a South African bank and other studies. S Afr J Sci. 2017;113(9/10), Art. #2016-0345, 7 pages. http://dx.doi.org/10.17159/sajs.2017/20160345

26. Govender P, Brooks MJ, Mathews AP. Cluster analysis for classification and forecasting of solar irradiance in Durban, South Africa. J Energy South Afr. 2018;29(2), Art. #4338, 12 pages. http://dx.doi.org/10.17159/2413-3051/2017/v29i2a4338

27. R Core Team. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2019. Available from: www.R-project.org

28. TIBCO Software Inc. Statistica version 13; 2018. Available from: http://tibco.com

AUTHOR:
Craig Pournara[1]

AFFILIATION:
[1]Wits School of Education, University of the Witwatersrand, Johannesburg, South Africa

CORRESPONDENCE TO:
Craig Pournara

EMAIL:
craig.pournara@wits.ac.za

DATES:
Received: 04 Apr.2020
Revised: 15 May 2020
Accepted: 18 May 2020
Published: 29 Sep. 2020

HOW TO CITE:
Pournara C. Grade 9 learners' algebra performance: Comparisons across quintiles, insights from errors and curriculum implications. S Afr J Sci. 2020;116(9/10), Art. #8125, 7 pages. https://doi.org/10.17159/sajs.2020/8125

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:
learner error, linear equations, didactic cut, integers, quintile 5 maths performance

FUNDING:
South African National Research Foundation (grant numbers 105901, 115261, 71218); First Rand Foundation

Grade 9 learners' algebra performance: Comparisons across quintiles, insights from errors and curriculum implications

It is well known that learners' performance in mathematics in under-resourced secondary schools in South Africa is poor. However, little is known about the performance of learners in top-performing well-resourced secondary schools apart from their good results in Grade 12. In this study, the performance of Grade 9 learners in top-performing quintile 5 (i.e. well-resourced) schools was compared with that of learners in quintile 1–3 (i.e. poorly resourced) schools using a 45-item test. While the quintile 5 learners obtained higher test scores, the performance pattern across the test items was very similar for both quintile groupings. A detailed error analysis of 89 quintile 5 learners' responses to equation items revealed difficulties in applying the standard equation-solving procedure, and in operating with negatives and subtraction, particularly on like terms. These problems may be related to a poorly conceived curriculum in the areas of integers and equations. It is recommended that Grade 8 and 9 teachers pay regular attention to all four arithmetic operations on integers. Furthermore, the teaching of equations should give greater attention to the cognitive shifts required in solving equations with letters on both sides of the equal sign.

Significance:

- Learner performance patterns on a test were similar for learners from top-performing quintile 5 schools and learners from lower quintile schools, although the quintile 5 learners obtained higher test scores.
- Quintile 5 learners' ability to solve linear equations correctly is substantially impacted by their difficulties in simplifying two algebraic terms to a single term, particularly when negatives and/or subtraction are involved.
- Particular aspects of the curriculum may partly be responsible for the difficulties learners experience with integers and equations.

Introduction

Research suggests that Grade 9 learners in quintile 5 schools, which are well resourced, are approximately 4 years ahead of their counterparts in poorly resourced quintile 1–3 schools.[1] This tells us about the relative performance of the two quintile groupings in South African schools and we know only too well about the poor mathematics performance of learners in under-resourced schools[2] but we still do not know much about what is happening at Grade 9 level in well-resourced quintile 5 schools. While many quintile 5 schools produce excellent results at Grade 12 level, research conducted on the mathematics performance of high performers writing the National Benchmark Tests[3] has shown that these learners have difficulties with apparently basic ideas such as percentage and inequalities. Many of these learners will come from quintile 5 and high-fee independent schools. Another indication that all is not well comes from informal discussions with heads of mathematics departments and teachers in quintile 5 schools who despair that many learners in Grades 8 and 9 are not performing at desired levels.

The Wits Maths Connect Secondary (WMCS) project is a research and development project at the University of the Witwatersrand. While our mandate is to focus on teacher professional development in lower quintile schools, we were curious to compare the performance of learners taught by teachers with whom we work, with the performance of Grade 9 learners in quintile 5 schools. We knew the quintile 5 learners would obtain higher marks but we wanted to compare *performance patterns* over the entire test, i.e. the trends in which items had a higher/lower number of correct responses. We also wanted to investigate learner errors and to compare these with previous findings of learners' performance in algebra in lower quintile schools.[4] As the ability to solve equations is fundamental for future success in mathematics, quintile 5 learners' responses to three linear equation items were investigated to gain insight into their fluency in solving equations and also into their fluency in algebraic manipulation. The research was framed by two questions:

- What similarities exist in the test performance patterns of quintile 5 and quintile 1–3 learners?
- What are the most common errors made by quintile 5 learners on linear equation items?

Literature review

Research on learners' approaches to solving linear equations and the errors they make goes back to the late 1980s.[5] There has been a resurgence in research on learner performance on linear equations and notions of equality in recent years.[6-8] One of the key findings of this accumulated research is that learners must be taught formal methods to solve equations of the form $ax+b=cx+d$ because their informal methods, which are adequate for equations of the form $ax+b=c$ break down for equations when there are letters on both sides and/or where there are two terms with letters on one side, e.g. $2x+5–x=4$. This breakdown (or discontinuity) has been referred to as the *didactic cut*[5] and the *cognitive gap*.[9] An equation such as $3x–4=11$, can be solved arithmetically by saying 'what multiplied by 3 and then subtract 4 gives me 11' or $3\times\square–4=11$? Clearly the solution is 5. However, this approach cannot

be applied to equations of the form $ax+b=cx+d$ and so learners must be taught to operate on the letters using inverse operations. The initial research on the didactic cut and cognitive gap involved learners who had not yet been taught formal procedures for solving equations. Research conducted with older learners who had already learned equation-solving procedures has challenged the existence of the didactic cut.[7]

Given that the research presented here also involves learners who have been taught procedures for solving equations, I shall rather use the notion of *epistemological obstacle*[10] in speaking about learners' difficulties in making the transition to formal methods for solving equations. An epistemological obstacle involves 'knowledge which functions well in a certain domain of activity and therefore becomes well-established, but then fails to work satisfactorily in another context where it malfunctions'[11]. Thus this notion of obstacle is concerned with a presence rather than an absence of knowledge. With reference to solving equations, the knowledge which has previously worked well refers to arithmetic approaches for solving equations. These methods need to be replaced with new knowledge for solving equations that have letters on both sides (or two terms with letters on one side).

The remainder of the literature review provides an overview of existing research on common errors in solving linear equations. This will provide the reader with the necessary background for the analysis which follows.

Approaches to and errors in solving equations

Kieran[12] identified seven approaches to solving equations, five of which are informal, including *undoing* or *working backwards* and *trial-and-error substitution*. She also distinguished two formal methods: *transposing of terms* (change side, change sign) and *performing the same operation on both sides*. The informal or *arithmetic* methods can be used for equations with letters on one side only while the formal or *algebraic* methods are necessary to solve efficiently equations with letters on both sides.

Four common errors have been identified in solving linear equations. Two of these are the *redistribution error* and *switching addends error*.[12] A redistribution error involves adding a term to one side but subtracting it from the other side. A switching addends error involves 'moving' a term across the equal sign without changing its sign. In this study, I refer to this as a *moving* error and I distinguish between moving constants and moving a letter-term. The *other inverse error*[13] occurs when learners use the incorrect inverse operation, e.g. given $5x=2$, a learner may subtract 5 from both sides instead of dividing by 5, giving $x=-3$ as the solution. Learners making the *familiar structure error*[14] 'force' their answer to fit the form $x=k$ by eliminating additional letters as necessary. For example, a learner who manipulates an equation to obtain $3x=12x$, might first divide by 3 to get $x=4x$ and then drop the letter on the right side and write $x=4$. I refer to this error as *familiar form* because it appears to be driven by learners' desire to produce a final answer of form $x=k$.

Meaning of the equal sign

Learners' conceptions of equality are clearly important in solving equations. Seminal research identified two different views of the equals sign: as a *do something signal* and as an indication of *equivalence.*[15,16] The former operational view is typically associated with unidirectional reasoning about equations and is frequently drawn on to solve equations of form $ax+b=c$. For example, as noted above in the case of $3x-4=11$, the learner reasons 'what multiplied by 3 and then subtract 4 gives me 11?' Here the learner treats the right side as the result of operations performed on the left side. The latter relational view is associated with solving equations of the form $ax+b=cx+d$. Research in the USA found that, across grade levels, learners who demonstrated a relational view of the equal sign, were better able to solve linear equations.[6] However, the authors note that despite learners' inadequate conceptions of equality, attention to the equal sign is typically not addressed in secondary school curricula in the USA. The same is true in the South African secondary curriculum.

Errors in operating on algebraic symbols

A fundamental component of early algebra involves making sense of new symbols and notation. In arithmetic $4+½=4\ ½$ but in algebra one cannot simply juxtapose the two symbols, i.e. $4+a\neq 4a$. In algebra, $a+b$ can be seen as the *process* of adding b to a as well as the resulting *object.*[17] The difficulties in making sense of the new notation explain, to some extent, why learners make errors when working with like and unlike terms, usually conjoining them to produce closure. I work with an expanded notion of conjoining which distinguishes *additive conjoining* from *subtractive conjoining* as two categories of errors that may involve like or unlike terms. While additive conjoining involves addition of positive terms, subtractive conjoining involves negatives and/or subtraction, e.g. $7-x=7x$; $3x-x=3$ and $-x+x=x$.

Errors with negatives

The minus symbol can be viewed as an *operator* (subtraction) or as a *sign* (negative). Hence when learners encounter an equation such as $2+3x=5-2x$, the $-2x$ can be seen as *subtracting* $2x$ or as *negative* $2x$. This duality of the minus symbol poses significant difficulty for learners. Equations involving negatives are more difficult because they are not easily modelled using a balance model.[18] Local research found that learners had greater difficulty in dealing with algebraic expressions when they involve negatives, either as sign or as operation.[4]

Research on subtraction and negatives has identified a range of errors associated with the minus symbol. For example, *right-to-left reasoning* involves subtracting a larger number from a smaller one[18], e.g. $4-7=3$ or $5x-7x=2x$. Overgeneralised integer rules may also lead to errors, e.g. if the multiplication rule is expressed as 'a minus *and* a minus gives a plus', this may lead to the expression $-2x-3x$ being simplified to $+5x$ because explicit attention is not given to the operation. The error of *detaching the minus sign* may occur when learners add numbers or terms with a leading negative[19], e.g. $-2+5=-7$ and $-3x+x=-4x$. In both cases learners detach the minus symbol and isolate it from the expression. They then perform the addition and re-attach the minus symbol to the answer.

Research design and methods

The research reported here stems from a larger study of learner performance. A one-hour test was administered to Grade 9 learners in late September/early October 2018. The first part of the analysis involved a comparison of the performance of these learners with that of pre-existing data from quintile 1–3 schools on the same test. The second part of the study involved further qualitative analysis of the quintile 5 data only with particular focus on learner errors in three test items.

Test instrument

The test consisted of 45 items dealing with number, algebra and function. Most items were typical curriculum items that learners would encounter in text books and tests, and they spanned Grades 7–9 content. The test had been piloted in 2016 with lower quintile schools but we were unsure how the items would perform with a quintile 5 sample. The comparative analysis of both quintile groupings deals with all 45 items. For reference, the broad content area of each question is given in Table 1.

Table 1: Test content areas and question numbers

Content area	Question numbers
Number	1, 2, 3, 4
Algebra	5, 6, 7, 9, 10
Pattern and function	8, 11, 12

Learner responses to each item were coded as *correct*, *incorrect* or *missing*, with no provision for partially correct responses. A learners' test score is simply a count of the number of fully correct responses. The response coding was led by members of the project team, with a group of research assistants. Coding and capturing of responses were moderated – approximately 15% and 20%, respectively. No errors were found in either process.

The error analysis and coding were conducted by the author alone. The focus was on three linear equation items, as shown in Table 2.

Table 2: Linear equation items

Item number	Item
Q9a	$3x-2=10$
Q9b	$3x-2=4+x$
Q9c	$2-3x=7-x$

Q9a contains a letter on the left side only and can easily be answered using arithmetic approaches. Q9b and Q9c contain letters on both sides but Q9c involves subtraction on both sides of the equation and is therefore more cognitively demanding than Q9b. These items and the associated learners' responses have potential to reveal evidence of learner difficulties in working with equations of the different forms. They also reveal learner errors when the combining of algebraic terms was not the main goal of the manipulations. Consequently, they have potential to reveal errors in solving equations as well as errors in simplifying algebraic expressions. This algebraic work would have been completed in the first half of the year and thus been examined by mid-year.

Sample

The sample from quintile 1–3 schools consisted of 1139 learners from 19 schools, taught by 25 teachers. Schools were selected because their mathematics teachers had completed a professional development course offered by the WMCS project in 2016 or 2017. The selected learners were taught by these teachers in 2018. The quintile 5 sample of 824 learners, taught by 22 teachers, was drawn from four quintile 5 secondary schools, all of which had an existing relationship with the WMCS project. They are top-performing schools in their respective districts and/or top feeder schools to the University of the Witwatersrand. The University's rankings for feeder schools are determined as a ratio of the number of applications to the number of enrolments from that school in a particular year. The large number of teachers is worth noting because it reduces the impact of individual teacher effects on the results.

The sub-sample for the error analysis consisted of 89 learners, across the four quintile 5 schools, who got Q9a correct but Q9b and Q9c incorrect. These criteria suggest possible evidence of an epistemological obstacle in solving more complex linear equations.

Ethical clearance was obtained from the University of the Witwatersrand ethics committee (H17/01/01) and the Gauteng Department of Education (M2017/400AA). All schools were assured that their identity would remain confidential and that no comparisons would be made between schools. Parents and learners were assured that the testing would not impact learners' marks and that they could withdraw at any point. They were also assured of the confidentiality of individual results.

Coding for error analysis

Learner responses were coded according to the approach used and the errors made. Because there are no interview data, it is difficult to infer the underlying reasoning informing learners' written responses. Coding was based on interpretations of what had been written, looking at changes between successive lines of a response together with individualised annotations which learners may have provided such as arrows indicating the moving of a term across the equal sign. While the approaches and errors are reported per item, I also compared each learner's responses to all three items, looking for similarities and differences that might assist in coding their errors.

I distinguished between algebraic and arithmetic approaches to solving the equations. For the purposes of this article, an algebraic approach involves manipulating expressions and operating on the letters. An arithmetic approach involves substitution of a possible solution or an undoing approach. For example, solving $3x–2=10$ by substitution might look as follows: $3(4)–2=10$. An undoing approach might be written as: $10+2\rightarrow12\div3\rightarrow4$

The error analysis was conducted on Q9b and Q9c. Drawing from other analyses of similar data[14,20], I distinguished three broad categories:

1. *Equation errors* – errors in applying inverse operations, collecting like terms and constants on opposites sides of the equal sign, and isolating the letter to determine the solution.
2. *Letter errors* – inappropriate or incorrect execution of operations on terms with letters.
3. *Numeric errors* – operations on constants where the outcome of the operation is incorrect. These are not reported here.

Each category was then sub-divided and errors were allocated specific codes, as described below. I included sub-codes for subtraction/negatives in each of the three categories. I assigned codes to each response based on the three broad categories as well as an 'other' category. It was possible for a single response to have multiple codes. I then dealt systematically with each category identifying sub-codes based both on the literature discussed above and on the data.

Equation error codes

Six sub-codes were identified for equation errors:

1. *Move term with letter* – a term involving a letter is moved unchanged across the equal sign
2. *Move constant* – a constant is moved unchanged across the equal sign
3. *Incorrect inverse* – additive inverse is applied when multiplicative inverse is required, or vice versa
4. *Divide binomial by monomial* – binomial is incorrectly divided by monomial to isolate letter on one side (see Figure 1a)
5. *Familiar format* – inappropriate manipulation of one/both sides of equation to force the form $x=k$ as the final line of the response. This code was only applied when comparing the last two lines of a response, as shown in Figure 1b.

In addition to the five codes above, I included an *incomplete* code for responses where the learner had not produced an answer in the form $x=k$ (see Figure 1c).

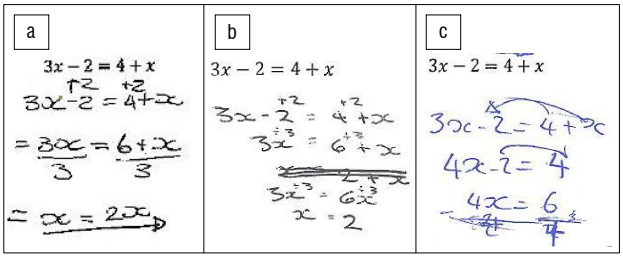


Figure 1: Equation errors: (a) binomial divided by monomial, (b) familiar format and (c) incomplete response.

Letter error codes

Letter errors were distinguished on two dimensions: those involving addition and/or positive terms, and those involving subtraction and/or negatives. The matrix in Table 3 provides examples of typical errors. The examples of subtraction/negatives with like terms include instances of right-to-left reasoning and detaching the minus sign.

Table 3: Examples of letter operation errors

	Like terms	Unlike terms
Addition/positives	$3x+x=3x$	$6+x=6x$ $4+x=5x$
Subtraction/negatives	$3x-x=3x$ $3x-x=3$ $-3x+x=-4x$ $x-3x=2x$	$3x-4=-x$ $7-x=7x$ $3-2x=1$

There were several instances of learners over-generalising exponential laws, e.g. $x+x=x^2$. These errors were separated from the addition-of-like-term errors shown above in order to determine the extent to which learners were still making typical conjoining errors that do not involve exponents. The errors involving exponents were coded as 'other'.

Analysis and results

The analysis is reported in two sections. I begin with the comparison of the overall performance and performance patterns of the quintile 5 group and the quintile 1–3 group. This is followed by the analysis of the responses of the quintile 5 sub-sample to the three equation items.

Overall performance and performance patterns

Table 4 shows that the mean score for the quintile 5 group (24.67) is more than 2.5 times the mean score of the quintile 1–3 group (9.34). This is to be expected and does not merit further discussion. However, a comparison of the performance patterns across the 45 items is of interest (see Figure 2). An obvious difference in the two graphs is that the quintile 5 group performed better than the other group on every item. Again, this is to be expected. More interesting is that the graphs have very similar shapes with peaks and dips in similar places. This suggests that learners across the quintiles find the same questions 'easy' and 'difficult'.

Table 4: Mean scores and standard deviations

Learner group	*N*	Mean	Standard deviation
Quintile 1–3	1139	9.34	6.64
Quintile 5	824	24.67	9.03

Given that the quintile 5 sample is drawn from top-performing schools, it may seem surprising that there were 21 items which fewer than 50% of learners answered correctly. There is a noticeable downward trend in the quintile 5 performance, interspersed with a few peaks. Better performance is generally associated with the numeric items in questions 1 to 4 which focus mainly on integers. Thereafter, most items involve algebra and this is where the downward trend becomes most noticeable. The three highest peaks from question 8 onwards are associated with numeric work: intercepts of the graph of a linear function (Q8a, Q8b); simple linear equation that can be solved without algebraic manipulation (Q9a); and a function machine with numeric inputs and outputs (Q11i, Q11ii). Thus the overall picture of quintile 5 performance is that learners have difficulty with algebraic work and functions. The downward trend is less obvious for the quintile 1–3 group because performance flattens from question 5 onwards, apart from the peaks which occur at similar places to those of the quintile 5 graph.

Three factors must be borne in mind when interpreting the lower than expected performance of the quintile 5 learners. Firstly, learners did not prepare for the test and so the scores merely provide a once-off measure on a particular day. Secondly, the data were collected in late September to early October and, in at least one school, teachers were still completing a section that was included in the test. Consequently, many learners may not yet have consolidated some of the test content. Thirdly, a learner's score indicates completely correct responses and so scores likely will under-estimate learners' knowledge of the test content. For example, a blank response and a response containing a minor slip in a calculation both count as zero.

Given the similarities in the performance pattern of the two groups, there is value in studying the errors made by the quintile 5 group because these will likely provide useful insights for both groups which may in turn lead to recommendations relevant to the teaching of algebra across all quintile schools.

Error analysis of responses to equations items

Table 5 provides a summary of learner performance on the equation items for both groups. In both groups there is a general decrease in performance from Q9a to Q9c with the percentage drop from Q9b to Q9c being larger. The drop in performance between Q9a and Q9b suggests that some learners experience some kind of obstacle in working with equations with letters on both sides. This appears to be exacerbated by the presence of additional negatives in Q9c, particularly given that the letters are being subtracted on both sides of the equation. It is also possible that performance on Q9c was lower because the solution is rational whereas the other two solutions are integers.

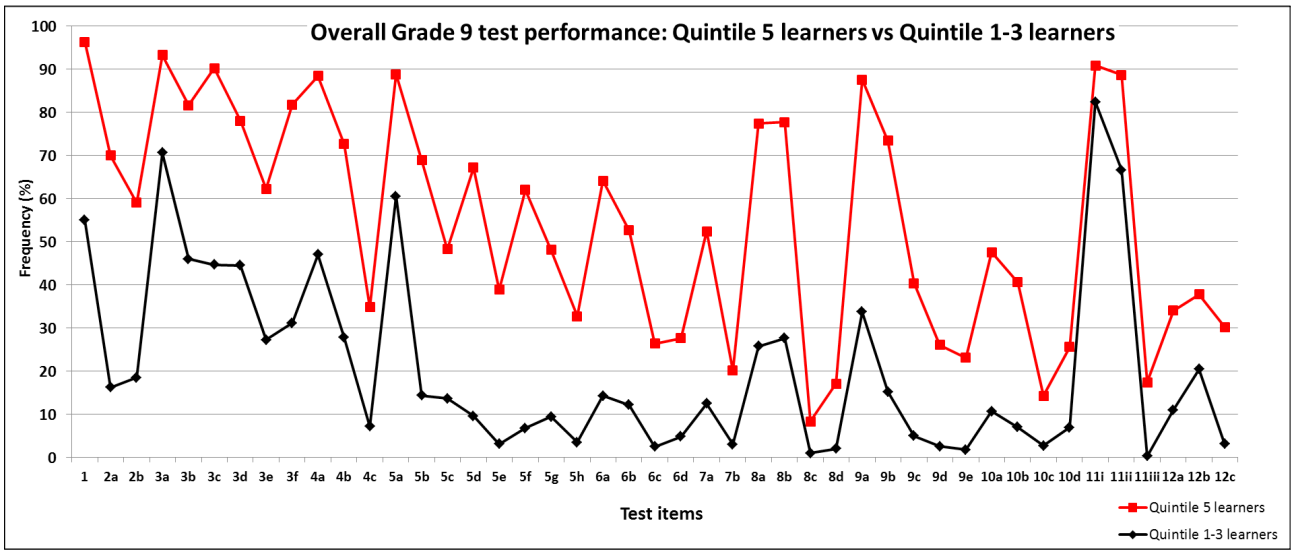


Figure 2: Learners' performance on all test items.

Table 5: Learner performance on linear equation items

Item number	Item	Quintile 5 ($N = 824$)	Quintile 1–3 ($N = 1139$)
Q9a	$3x-2=10$	87.6	33.80
Q9b	$3x-2=4+x$	73.4	15.28
Q9c	$2-3x=7-x$	40.4	5.00

Learners' approaches to solving equations

The analysis from here focuses on the quintile 5 sub-sample. Learners' approaches to Q9a in contrast to their approaches to Q9b and Q9c were of particular interest because this might reveal the extent to which they have overcome the epistemological obstacle described earlier. Table 6 shows that the vast majority used algebraic approaches for Q9a. Of greater interest, however, are the learners who used arithmetic approaches.

Table 6: Approaches to Q9a

Approach	Number of responses	Percentage
Algebraic	63	73.3
Arithmetic Substitution	15	17.4
Undoing	3	3.5
No evidence	5	5.8
Total	**86**	**100**

For those using arithmetic approaches to correctly answer Q9a, it was important to investigate their approaches to Q9b and Q9c: did they shift to an algebraic approach or did they still attempt an arithmetic approach? Eighteen learners successfully used arithmetic approaches for Q9a, either substitution or an undoing approach. Only two of these learners attempted an algebraic approach for Q9b and Q9c.

Of the 15 learners who used a substitution approach for Q9a, 11 used the same approach for Q9b and 9 used it for Q9c. Of the three learners who used an undoing approach for Q9a, two attempted this approach for Q9b and Q9c while the third learner approached both items algebraically. There were five learners who only provided answers and so their responses were coded as 'no evidence'.

In summary, most learners either used all arithmetic approaches or all algebraic approaches. Those adopting arithmetic approaches for all three items constituted 23.8% of the sub-sample. It appears that these learners have not yet recognised the need to reject an inadequate method and replace it with a procedure that makes use of inverse operations. These learners are not operating on the terms with letters in order to solve the equations. It is worth noting that the mean test score for these 15 learners was 14.2 (31.6%) which suggests that they lack algebraic fluency more generally.

I now shift to a discussion of the errors made by the learners who used algebraic approaches for Q9b and Q9c.

Errors in equation operations

A summary of the equation errors is presented in Table 7. A total of 112 equation errors were coded across Q9b and Q9c.

Table 7: Equation errors for Q9b and Q9c

Equation errors	Q9b	Q9c
Move term with letter	14	7
Move constant	6	11
Incorrect inverse	3	2
Divide binomial by monomial	3	3
Force familiar format for solution	13	14
Incomplete	6	13
Other	6	11
Total	**51**	**61**

Not surprisingly, the most common error related to the incorrect use of additive inverses, with 38 errors involving 'moving' a letter/constant across the equal sign without changing sign. It is worth noting that only four learners made 'moving errors' with both numbers and letters. There were five instances where learners used the incorrect inverse, typically subtracting a coefficient instead of dividing. In a small number of instances learners did not collect like terms with letters and so divided a binomial on one side of the equation by the coefficient of the variable on the other side. This typically led to other errors: either the learner performed the division on the numbers only and then conjoined the remaining number and letter (see Figure 1a), or dropped the letter from the expression.

There were a surprising number of responses (27) where learners manipulated the equation to force a familiar format, i.e. $x=k$, as the final line of the response. This took different forms, including dropping/ignoring 'unwanted' letters, e.g. the equation $3x=6x$ became $x=2$.

The high prevalence of *incomplete* responses (19) is closely linked to the familiar format error. In many of these cases, learners did not manipulate their equations to produce a familiar format. Instead they stopped with forms such as $ax=b$, $x=ax$ or $ax/a=b/a$.

The *other* equation errors included converting an equation to an expression, incorrectly combining terms to generate quadratic forms, and various manipulations that did not maintain equality.

Errors in letter operations

A total of 98 letter errors were coded for Q9b and Q9c. As described above, I disaggregated conjoining errors to distinguish (1) additive conjoining (addition/positives) from subtractive conjoining (subtraction/negatives); and (2) operating on like terms from operating on unlike terms. This makes it possible to see more clearly where the majority of errors occurs.

Table 8: Errors in letter operations

	Like terms	Unlike terms	Total
Addition/positives	2	6	**8**
Subtraction/negatives	42	13	**55**
Total	**44**	**19**	**63**

As can be seen in Table 8, only 8 errors involved additive conjoining, irrespective of whether learners were operating on like or unlike terms. By contrast, 55 errors involved subtraction/negatives and 42 of these involved *like* terms. This may be surprising in the light of prior research on conjoining. However, assuming most learners in this sample know we cannot combine unlike terms but can combine like terms, then it is not surprising that the majority of errors occurs when combining like terms. Typical errors included detaching the negative, e.g. $-3x+x=-4x$ and dropping the letter, e.g. $3x-x=3$. There was a surprisingly high number of responses giving $3x-x=3x$. One interpretation is that learners consider only the visible coefficients. In effect they are treating x as $0x$ rather than $1x$. If they separate the numbers and letters, they may reason $3-0=3$ and then append the x to obtain $3x$. However, interviews are necessary to investigate this further.

In addition to the above errors, there were 16 instances in Q9c where learners dropped the negative sign from one line to the next. For example, $-3x$ became $3x$ in the following line. There were also 19 letter errors coded as *other.* These mostly involved over-generalisation of the addition law of exponents which typically led to further errors in attempting to solve an equation that was no longer linear.

Discussion and implications

The overall Grade 9 learner performance on a test of number, algebra and function covering selected Grade 7–9 content was disappointing for both quintile groupings. However, both groupings displayed similar performance patterns. Notwithstanding the caveats mentioned above, a mean score of 54.8% for the quintile 5 group indicates that even towards the end of Grade 9 there are learners in top-performing schools who still have difficulty manipulating algebraic symbols.

Fundamental to the notion of epistemological obstacle in the context of solving equations is that learners accept the need to replace their inappropriate arithmetic approaches with an algebraic approach. Although most learners attempted algebraic approaches for all three items, approximately 24% of the sub-sample used only arithmetic approaches. This may also suggest they do not have a relational view of the equal sign. Furthermore, the error analysis shows that, on average, each learner in the sub-sample made more than one equation error and one letter error across Q9b and Q9c.

The error analysis reveals that errors made by quintile 5 learners in solving linear equations with letters on both sides stem more from difficulties in manipulating algebraic expressions and dealing with negativity than in executing the standard procedure for solving equations. Of the errors reported here, 92 (43.8%) errors relate to negatives/subtraction in some way. Furthermore, nearly half (45.6%) of these negativity errors involved the incorrect simplification of two like terms to a single term. While these findings confirm some of what we found in a previous study on learners' algebra performance in lower quintile schools[4], the insight, at least for quintile 5 schools – that difficulties with negatives and subtraction are more common with *like* terms – is a new empirical finding, although not necessarily surprising.

Implications for curriculum and teaching

From the above findings there are two clear implications for the curriculum and four implications for teaching.

An analysis of the Senior Phase Mathematics curriculum[21] suggests that two problems highlighted in this study may have their roots in the curriculum itself. For example, the content of integers is split over Grades 7 and 8, a split exacerbated by the move from primary to secondary school. The Grade 8 curriculum assumes learners come with knowledge of adding and subtracting integers, that this merely requires revision and that teachers should focus on multiplication and division of integers in Grade 8. While teachers may ignore this 'advice' in their own classrooms, the official teaching support materials such as annual teaching plans, scripted lesson plans and learner workbooks will follow the curriculum closely and thus fall prey to the poorly conceived plans for teaching and learning integers. The evidence from this study and prior research shows clearly that all aspects of integers need detailed attention in Grade 8.

A similar problem arises with equations. There is considerable focus on solving equations by inspection and insufficient attention to formal equation operations in Grade 8. Also, there is no explicit recognition of the importance of attending to equations with letters on both sides. By Grade 9, it is assumed that learners have mastered this work and can move on to more complex linear examples as well as quadratic and exponential equations. Given this breadth of equation types, teachers may overlook the need to deal with simple linear equations with letters on both sides, in the rush to cover the other types. The curriculum needs to foreground the cognitive shifts in moving from equations with a letter on one side to equations with letters on both sides, with additional time allocated to consolidate these procedures, thus supporting learners to navigate and overcome the epistemological obstacle they encounter when they have to operate on the letter in solving equations. The overburdened curriculum could be eased by removing quadratic and exponential equations from Grade 9 as they are dealt with in detail in later years.

Implications for teaching follow closely from the curriculum implications. Firstly, teachers should pay explicit attention to helping learners develop an equivalence view of the equal sign, even in Grades 8 and 9. Without an equivalence view, learners will continue to have difficulty in solving equations of all kinds. Secondly, Grade 8 teachers should pay attention to all four arithmetic operations on integers, with particular attention to subtraction. Attention to fluency with negative numbers should continue into Grade 9. Thirdly, continual attention must be given to fluency in algebraic manipulation, particularly with examples involving subtraction and negatives. This study suggests that such a focus will improve learners' performance on equation solving. Fourthly, teachers need to appreciate the cognitive shift necessary to solve equations with letters on both sides and take time to deal with the case of $ax+b=cx+d$. They should also include equations with more than two terms on each side, e.g. $4-2x+3=3x+1-x$. This provides practice in algebraic simplification as well as in performing inverse operations.

Conclusion

This study shows that difficulties with introductory algebra are not restricted to learners in lower quintile schools. Furthermore, it makes three important contributions. Firstly, there are similarities in the performance patterns of Grade 9 learners across quintiles on a test of number, algebra and function spanning content of Grades 7 to 9. Secondly, it reveals and confirms learners' specific difficulties in working with negatives and subtraction in relation to algebra. Thirdly, it highlights the specific insight that while few learners were making errors with addition of like and unlike terms, there was a proliferation of errors in working with like terms and negatives.

While many learners in quintile 5 schools overcome these difficulties and perform well in mathematics by Grade 12, the same cannot be said for the majority of learners in lower quintile schools. The curriculum recommendations proposed above suggest that specific curriculum changes are necessary in the topics of integers and equations. These may help to address the ways in which the curriculum contributes to learner difficulties with negative numbers and aspects of algebra. The recommendations for teaching address similar issues. However, opportunity for teachers to implement the recommendations requires some flexibility in curriculum pacing to address learners' errors and backlogs.

Acknowledgements

This work is based on research supported by the National Research Foundation (NRF) of South Africa (grant numbers 105901, 115261, 71218) and the First Rand Foundation (FRF). All opinions, findings and conclusions or recommendations are that of the author and the NRF and FRF accept no liability whatsoever in this regard. I am grateful to my colleagues in the WMCS project for many hours of productive discussions on our research and development work, particularly in relation to the teaching and learning of algebra.

Competing interests

I declare that there are no competing interests.

References

1. Spaull N, Kotze J. Starting behind and staying behind in South Africa: The case of insurmountable learning deficits in mathematics. Int J Educ Dev. 2015;41:13–24. https://doi.org/10.1016/j.ijedudev.2015.01.002
2. Reddy V, Van der Berg S, Janse van Rensburg D, Taylor S. Educational outcomes: Pathways and performance in South African high schools. S Afr J Sci. 2012;108(3/4), Art. #620, 8 pages. https://doi.org/10.4102/sajs.v108i3/4.620
3. Bohlmann C, Frith V, Prince R. Mathematical errors made by high performing candidates writing the National Benchmark Tests. Pythagoras. 2017;38(1), Art. #292. https://doi.org/10.4102/pythagoras.v38i1.292
4. Pournara C, Hodgen J, Sanders Y, Adler J. Learners' errors in secondary algebra: Insights from tracking a cohort from Grade 9 to Grade 11 on a diagnostic algebra test. Pythagoras. 2016;37(1), Art. #334. https://doi.org/10.4102/pythagoras.v37i1.334
5. Filloy E, Rojano T. Solving equations: The transition from arithmetic to algebra. For the Learning of Mathematics. 1989;9(2):19–25.
6. Knuth E, Stephens A, McNeil N, Alibali M. Does understanding the equal sign matter? Evidence from solving equations. J Res Math Educ. 2006;37(4):297–312.
7. Lima R, Healy L. The didactic cut in equation solving or a gap between the embodied and the symbolic mathematical worlds? In: Pinto M, editor. Proceedings of the 34th Conference of the International Group for the Psychology of Mathematics Education. Belo Horizonte, Brazil: PME; 2010. p. 353–360.
8. Roberts A, Le Roux K. A commognitive perspective on Grade 8 and Grade 9 learner thinking about linear equations. Pythagoras. 2018;40(1), Art. #438. https://doi.org/10.4102/pythagoras.v40i1.438
9. Herscovics N, Linchevski L. A cognitive gap between arithmetic and algebra. Educ Stud Math. 1994;27:59–78. https://doi.org/10.1007/bf01284528
10. Brousseau G. Theory of didactical situations in mathematics. Dordrecht: Kluwer; 1997.
11. Cornu B. Limits. In: Tall D, editor. Advanced mathematical thinking. Dordrecht: Kluwer; 1991. p. 153–166.
12. Kieran C. The learning and teaching of school algebra. In: Grouws D, editor. Handbook of research on mathematics teaching and learning. New York: Macmillan; 1992. p. 390–419.
13. Hall R. An analysis of errors made in the solution of simple linear equations. Phil Math Educ J. 2002;15(1):1–67.
14. Sanders Y. Learners' performance in arithmetic equivalences and linear equations Johannesburg: University of the Witwatersrand, 2017.
15. Behr M, Erlwanger S, Nichols E. How children view the equals sign. Math Teach. 1980;92:3–15.
16. Kieran C. Concepts associated with the equality symbol. Educ Stud Math. 1981;12:317–326.
17. Sfard A. On the dual nature of mathematical conceptions: Reflections on processes and objects as different sides of the same coin. Educ Stud Math. 1991;22:1–36. https://doi.org/10.1007/bf00302715
18. Vlassis J. Making sense of the minus sign or becoming flexible in 'negativity'. Learn Instruct. 2004;14(5):469–484. https://doi.org/10.1016/j.learninstruc.2004.06.012
19. Linchevski L, Herscovics N. Crossing the cognitive gap between arithmetic and algebra: Operating on the unknown in the context of equations. Educ Stud Math. 1996;30:39–65. https://doi.org/10.1007/bf00163752
20. Halley A. Linear equations: Grade 9 learners' progress and their difficulties over time. Johannesburg: University of the Witwatersrand; 2019.
21. South African Department of Basic Education (DBE). Curriculum and Assessment Policy Statement Grades 7-9 Mathematics. Pretoria: DBE; 2011.

Newly identified hominin trackways from the Cape south coast of South Africa

AUTHORS:
Charles W. Helm[1]
Martin G. Lockley[2]
Hayley C. Cawthra[1,3]
Jan C. De Vynck[1]
Mark G. Dixon[1]
Carina J.Z. Helm[1]
Guy H.H. Thesen[1]

AFFILIATIONS:
[1]African Centre for Coastal Palaeoscience, Nelson Mandela University, Port Elizabeth, South Africa
[2]Dinosaur Trackers Research Group, University of Colorado, Denver, Colorado, USA
[3]Geophysics and Remote Sensing Unit, Council for Geoscience, Cape Town, South Africa

CORRESPONDENCE TO:
Charles Helm

EMAIL:
helm.c.w@gmail.com

DATES:
Received: 12 Apr. 2020
Revised: 31 May 2020
Accepted: 17 June 2020
Published: 29 Sep. 2020

HOW TO CITE:
Helm CW, Lockley MG, Cawthra HC, De Vynck JC, Dixon MG, Helm CJZ, Thesen GHH. Newly identified hominin trackways from the Cape south coast of South Africa. S Afr J Sci. 2020;116(9/10), Art. #8156, 13 pages. https://doi.org/10.17159/sajs.2020/8156

ARTICLE INCLUDES:
☒ Peer review
☒ Supplementary material

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

EDITOR:
Margaret Avery

KEYWORDS:
Homo sapiens, fossil tracks, aeolianites, southern Africa, Pleistocene

FUNDING:
None

Three new Pleistocene hominin tracksites have been identified on the Cape south coast of South Africa, one in the Garden Route National Park and two in the Goukamma Nature Reserve, probably dating to Marine Isotope Stage 5. As a result, southern Africa now boasts six hominin tracksites, which are collectively the oldest sites in the world that are attributed to *Homo sapiens*. The tracks were registered on dune surfaces, now preserved in aeolianites. Tracks of varying size were present at two sites, indicating the presence of more than one trackmaker, and raising the possibility of family groups. A total of 18 and 32 tracks were recorded at these two sites, respectively. Ammoglyphs were present at one site. Although track quality was not optimal, and large aeolianite surface exposures are rare in the region, these sites prove the capacity of coastal aeolianites to yield such discoveries, and they contribute to what remains a sparse global hominin track record. It is evident that hominin tracks are more common in southern Africa than was previously supposed.

Significance:

- Three new Pleistocene hominin trackways have been identified on the Cape south coast, bringing the number of known fossil hominin tracksites in southern Africa to six.
- The tracks were all registered on dune surfaces, now preserved as aeolianites.
- These are the six oldest tracksites in the world that are attributed to *Homo sapiens*.
- Hominin tracks are more common in southern Africa than was previously supposed.

Introduction

The Cape south coast of South Africa has been shown to be of pivotal importance in the origin of cognitively modern humans in the Middle Stone Age, with sites such as Blombos Cave[1], Pinnacle Point[2], and Klasies River[3] achieving international renown. Examples of the early emergence of modern human behaviour include adornment through the use of ochre[1,4], creation of jewellery[5], development of microlithic technology[6], heat treatment of stone tools[7], manufacture of bone tools[8], and the creation of abstract symbols, both as petroglyphs[9,10] and pictographs[11]. Marean[2], documenting the first known systematic exploitation of seafood by humans, postulated that the Cape south coast formed a refugium that facilitated human survival during the harsh climate of Marine Isotope Stage (MIS) 6.

Fortuitously, the surfaces of unconsolidated dunes on which those early humans trod and travelled during the Pleistocene have the capacity to be preserved on the current coastline as aeolianites (cemented dune deposits), and to provide a record of events that transpired on them. Ichnology, the study of tracks and traces, therefore has the potential to complement this corpus of research. Indeed, our Cape south coast ichnology project has documented more than 250 vertebrate tracksites in our study area (Figure 1), which extends for 350 km from the town of Arniston in the west to the Robberg peninsula in the east.[12] For example, the presence of giraffe[13], crocodiles[14], and breeding sea turtles[15] was not suspected from the skeletal record, and has only been established through the presence of their tracks.

Ichnological evidence of the human presence on this coast was confirmed through description of a hominin tracksite at Brenton-on-Sea.[16] This discovery added to the sparse number of reported Pleistocene southern African hominin tracksites, namely Langebaan on the west coast and Nahoon on the southeast coast.[17] While the Langebaan tracks have not been universally accepted as having been made by humans[18], for the purposes of this article we follow the consensus that they are of probable hominin origin. As a result, southern Africa currently lays claim to the only hominin tracksites in the world older than 46 ka that have been attributed to *Homo sapiens*.[16,18-21]

In our continuing exploration of the Cape south coast we have identified further tracksites, which we are comfortable in ascribing to hominin trackmakers. The purpose of this article is to describe these newly identified ichnosites, to consider criteria for the identification of hominin tracks in this region in softer substrates, and to consider the implications of these discoveries. We confine ourselves here to tracks of unshod humans; the possibility of humans sometimes using footwear in this region during the Middle Stone Age will be discussed elsewhere.

Background

The global hominin track record

Comprehensive reviews of global hominin tracksites and hominin ichnotaxonomy are found in Lockley et al.[19,20,22], Kim et al.[23], and Bennett and Morse[18]. Lockley et al.[19] listed 63 sites, and Bennett and Morse[18] listed 44 sites. Further sites have been reported subsequent to the publication of these reviews.[16,24,25] Compared with the three southern African sites thus far reported, there are only six generally accepted older hominin tracksites in the world[18-20], and none of these are attributed to *Homo sapiens*. Thus southern Africa has to date proven to be the region in which to search for Middle Stone Age tracks of our species.

While the southern African sites (including the new sites we describe here) all occur on aeolianite surfaces, this is not a common phenomenon from a global perspective. Solidified volcanic ash deposits or undisturbed cave floor deposits are not only commoner, but also have the capacity to preserve track morphology in greater detail than do the southern African aeolianites. Belvedere and Farlow[26] proposed a four-point scale (0–1–2–3) for quantifying the preservation quality of vertebrate tracks. Tracks registered and preserved in dune facies usually will not score more than 2 on this scale. Another feature of the global record is that the majority of hominin tracksites are preserved in epirelief (i.e. as natural impressions on the original surface), while in southern Africa these sites are mostly preserved in hyporelief (i.e. as natural casts on the infill surface).

The southern African hominin track record

The three previously reported southern African sites (Figure 1), in order of their initial description, are:

1. The Nahoon site, situated in South Africa's Eastern Cape province, and comprising three hominin tracks in hyporelief, was identified in 1964.[17,27] The tracks occur in the Nahoon Formation of the Algoa Group.[28] Soon after its discovery, the in-situ aeolianite slab containing the tracks collapsed. However, the tracks were recovered, and are housed in the East London Museum (South Africa). They have been dated through optical stimulation luminescence (OSL) to ~124 ka.[29]
2. The Langebaan tracksite, situated on the Cape west coast in South Africa's Western Cape Province, and comprising three probable hominin tracks in epirelief, was identified in 1995.[17,30] The tracks occur in the Langebaan Formation of the Sandveld Group.[31] The tracks were recovered and are housed in the Iziko South Africa Museum, Cape Town. They have been dated through OSL to ~117 ka.[17]
3. The Brenton-on-Sea site, comprising as many as 40 tracks in hyporelief and cross-section on the ceiling and walls of a small coastal cave, was identified in 2016 and reported on in 2018.[16] The tracks occur in aeolianites in the Waenhuiskrans Formation of the Bredasdorp Group.[32] The tracks were made by a party of humans moving rapidly down a dune slope.[16] Short trackways were evident. An age estimate of 90 ka was obtained using carbonate diagenesis and stratigraphic correlation to nearby dated sites.[16] The tracks were not manually recoverable, and a digital record was obtained through photogrammetry. OSL results are awaited.

Tuttle[33] developed criteria for hominin track identification:

- The hallux (digit I) is aligned with the four lateral toes (digits II–V), which are short and straight.
- The tip of the hallux is bulbous, not tapered.
- The tips of the hallux and adjacent second and third toes do not project markedly beyond one another.
- A prominent medial longitudinal arch is evident.

Morse et al.[34], in reporting on Holocene tracks (including human tracks) from Namibia, examined the influence of different substrates on footprint morphology. Anticipating the possible identification of further southern African sites, and appreciating that Tuttle's criteria[33] applied to near-ideal substrates and track preservation, Helm et al.[21] reviewed described sites from the region and considered putative sites, developing criteria on which to base the investigation of future discoveries.

In a further development, Helm et al.[35] reported on evidence that hominins left more than just tracks on these surfaces, and that patterns indicating a hominin 'signature' could be identified on aeolianite surfaces. These patterns could represent the creation of palaeo-art or abstract motifs, messaging or foraging. The term 'ammoglyph' was coined to describe such findings. One site included possible human knee impressions[35], adding to a very sparse global record of such features[36,37].

Next described was the first example on the Cape south coast of Middle Stone Age tools embedded in a Pleistocene track-bearing palaeo-surface, containing tracks of crocodiles, birds and mammals.[14] The tracksites we describe here therefore need to be considered in the broader context of features related to human activity on Pleistocene aeolianite surfaces on the Cape south coast, including ammoglyphs and Middle Stone Age tools, and the possibility of shod human tracks.

Geological context

Pleistocene sediments on the Cape south coast form part of the Bredasdorp Group. Within this Group, Pleistocene aeolianites, cemented palaeodunes and interdune areas of calcarenite[38], form part of the Waenhuiskrans Formation[32]. We are confident that the tracksites described here occur in situ in aeolianites, and not in marine deposits of the Klein Brak Formation.[39] These Pleistocene deposits have been dated to between 400 ka and 36 ka through OSL[40-45], with dated results indicating several MIS 5 deposits.

Aeolianites extend intermittently along much of South Africa's present-day coastline, and are well exposed in embayments on or near the shoreline.[46] They are composed of medium- to fine-grained sand with a high carbonate content derived from marine shell fragments. They lithify as a result of the downward percolation of rainwater in the meteoric diagenetic zone; the carbonate shell component is mobilised and re-deposited as interstitial cement within the sandstone matrix.[47]

The sites described here occur within or close to the Wilderness Embayment.[44,48,49] Here the Pleistocene stratigraphy consists mostly of

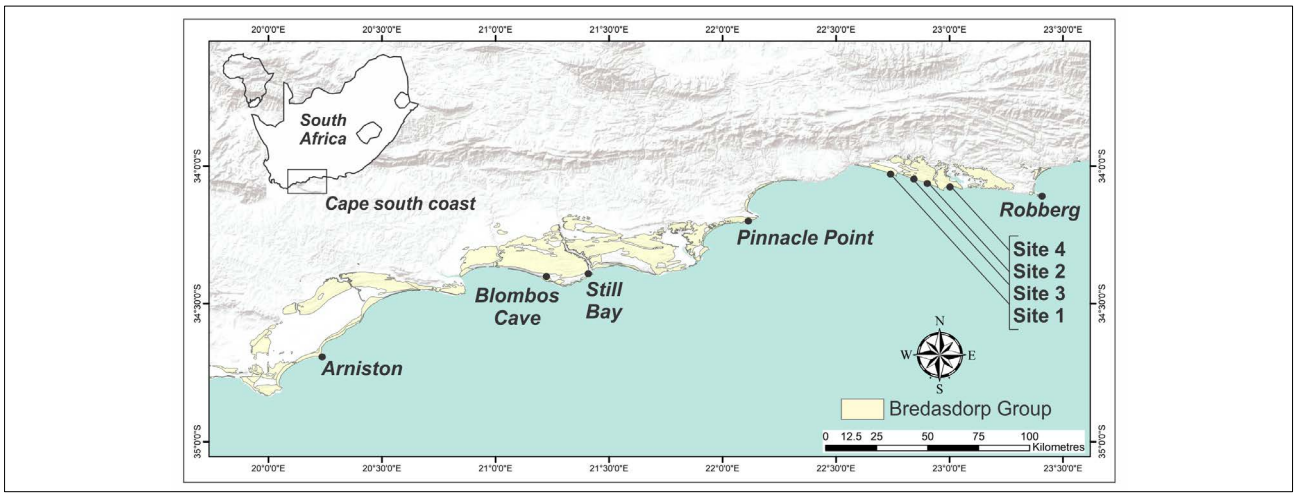


Figure 1: Map of South Africa and the Cape south coast, showing places mentioned and the locations of Site 1 – Site 4, and the extent of outcrops of the Bredasdorp Group.

MIS 6 and MIS 5 deposits, draped by an unconsolidated unit of modern dunes.[44] OSL dates from this region range from 148 ka ± 10 ka to 79 ka ± 9 ka.[44]

The distance of the tracks from the coastline at the time they were registered is related to Pleistocene sea-level oscillations and the periodic exposure of the Palaeo-Agulhas Plain, as a result of global temperature changes. For example, around 91 ka, during MIS 5c, sea levels were as much as 45 m lower than at present, and the coastline in this area was up to 60 km seaward of the present-day coast.[50-52] In contrast, around 126 ka, at the time of the peak MIS 5e high-stand[53], sea levels were 6–8 m higher than present[41,54]. Because Quaternary tectonic activity is considered minimal along the Cape south coast[55], in-situ strata lie at or close to their original angles of deposition, often at the angle of repose of wind-blown sand.

Less than a quarter of the Cape south coast comprises Pleistocene exposures. The remainder consists of Palaeozoic quartzite, sandstone and shale exposures of the Cape Supergroup, Precambrian granite exposures of the Cape Granite Suite, and expanses of beach and unconsolidated Holocene beach and dune sediments.[56]

Tracksites, once exposed, are subjected to erosive forces of water and wind, including wave action. Consequently, there is often a relatively short window in which to document such sites before their quality deteriorates or they are destroyed entirely.

Methods

Global Positioning System readings were taken using a handheld device. Locality data were reposited with the African Centre for Coastal Palaeoscience at Nelson Mandela University, South Africa, to be made available to bona fide researchers upon request. Tracksites were interpreted in the field through correlation to dated deposits, and samples were taken for OSL dating.

Where appropriate, track length, track width, track depth, pace length, stride length, and thickness of foresets were measured.[57] Results were recorded in centimetres. Dip and strike measurements were recorded. Where feasible, a 50 cm x 50 cm grid system was employed to create a sketch map of the track-bearing surfaces, along with a numbering system for the tracks. Trackway maps were developed using Adobe Illustrator (version 23.0.3) and Microsoft PowerPoint (version 1908).

Photographs were taken, including images for photogrammetry.[58] Photogrammetry 3D models were generated with Agisoft MetaShape Professional (v. 1.0.4) using an Olympus TG-5 camera (focal length 4.5 mm; resolution 4000 x 3000; pixel size 1.56 x 1.56 um). The final images were rendered using CloudCompare (v2.6.3.beta); details are provided in the relevant figure captions.

Results

Site 1: Garden Route National Park

Site 1 lies within the coastal section of the Garden Route National Park, between the towns of Wilderness and Sedgefield. The tracks occur on the ceiling of a cave at the foot of aeolianite cliffs. We generated a trackway map of this surface (Figure 2). The strike of this surface is 100°, with a gentle dip of 4°. This cave is situated near the high water mark (Figure 3a). Its interior and ceiling are buffeted by waves during high tides and storm surges.

As many as 18 tracks occur as natural casts (i.e. in hyporelief) on this ceiling, although some of these have indistinct margins or are not accessible for analysis. For this reason, we have only assigned identifier labels to seven of the tracks (Track A – Track G). The tracks have a compound aspect, i.e. the stratigraphically lower expressions are rounded oval 'casts' (convex hyporeliefs) resembling a boat hull, and interpreted as transmitted tracks impacting layers a few laminae below where the trackmakers made direct 'true track' contact with the substrate.[59] Where these transmitted track surfaces have been eroded into, it is possible to observe the cross-sections of the nested, convex-down laminae, just above the more widely exposed surface revealing the transmitted track casts in hyporelief. Thus the generic term 'track' can refer to the stacked sequence in which the 'true track' is nested in, and expressed by, its underlying transmitted track features.

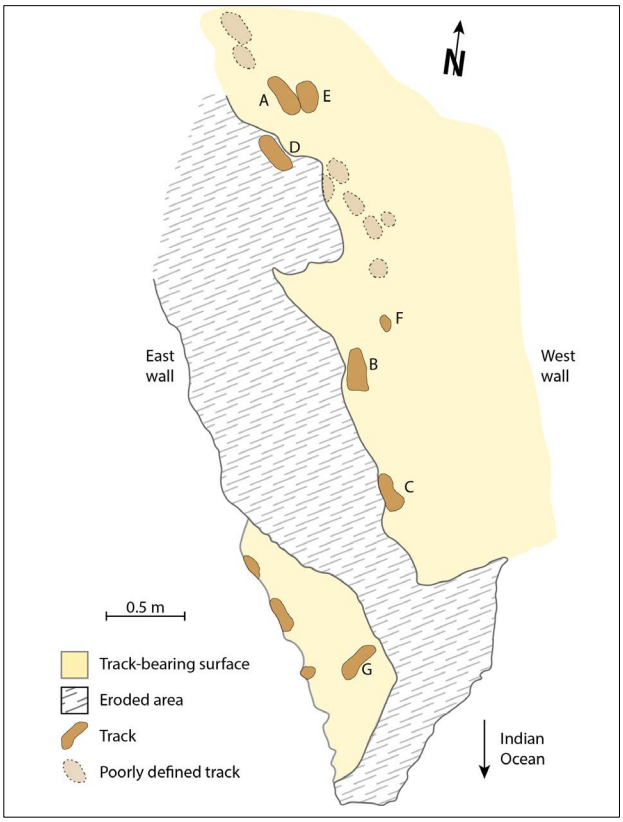


Figure 2: Trackway map of Site 1.

The cave has a sandy floor; the vertical distance from floor to ceiling is 190 cm at the dripline at the seaward end. From here it extends into the cliffs in a northerly direction, and the vertical distance between floor and ceiling decreases to a minimum of 75 cm (Figure 3b). From this point it continues as a tunnel in a more northwesterly direction to a northwestern entrance, with a vertical distance between floor and ceiling of 150 cm at this entrance. The total length of this cave/tunnel is 8 m. All the tracks occur in the seaward (southern) portion.

Two further track-bearing surfaces are preserved in epirelief below the main layer. One surface lies 43 cm below the main layer, and contains ~24 small, unidentifiable tracks on a rippled surface. The second lies a further 8 cm lower in the section, and contains 16 small- and medium-sized bovid tracks. Still lower in the section are heavily bioturbated layers. The tracks of the main track-bearing surface in the deeper portion of the cave can only be accessed with some difficulty via a 30 cm gap between bedding planes.

The site was initially identified in 2013, but only analysed in detail in 2019. While photographs confirm that there was minimal degradation of the tracks between 2013 and 2019, a significant interval change occurred over a 6-week period in 2019 following storm surges, resulting in the partial loss of one of the better-preserved tracks (Track A). Further subsequent degradation of the surface has occurred (Figure 3c, 3d).

With one exception (described below) the tracks have a similar south-southwest orientation of ~222°. The main trackway appears very straight, with a narrow straddle (Figure 4a). While some of these tracks have indistinct margins, the three best-preserved tracks exhibit the following dimensions:

- Track A: 24 cm long, 11 cm wide, maximum depth 6.5 cm
- Track B: 26 cm long, 12 cm wide, maximum depth of 7.5 cm
- Track C: 24 cm long, 11 cm wide, depth not measurable

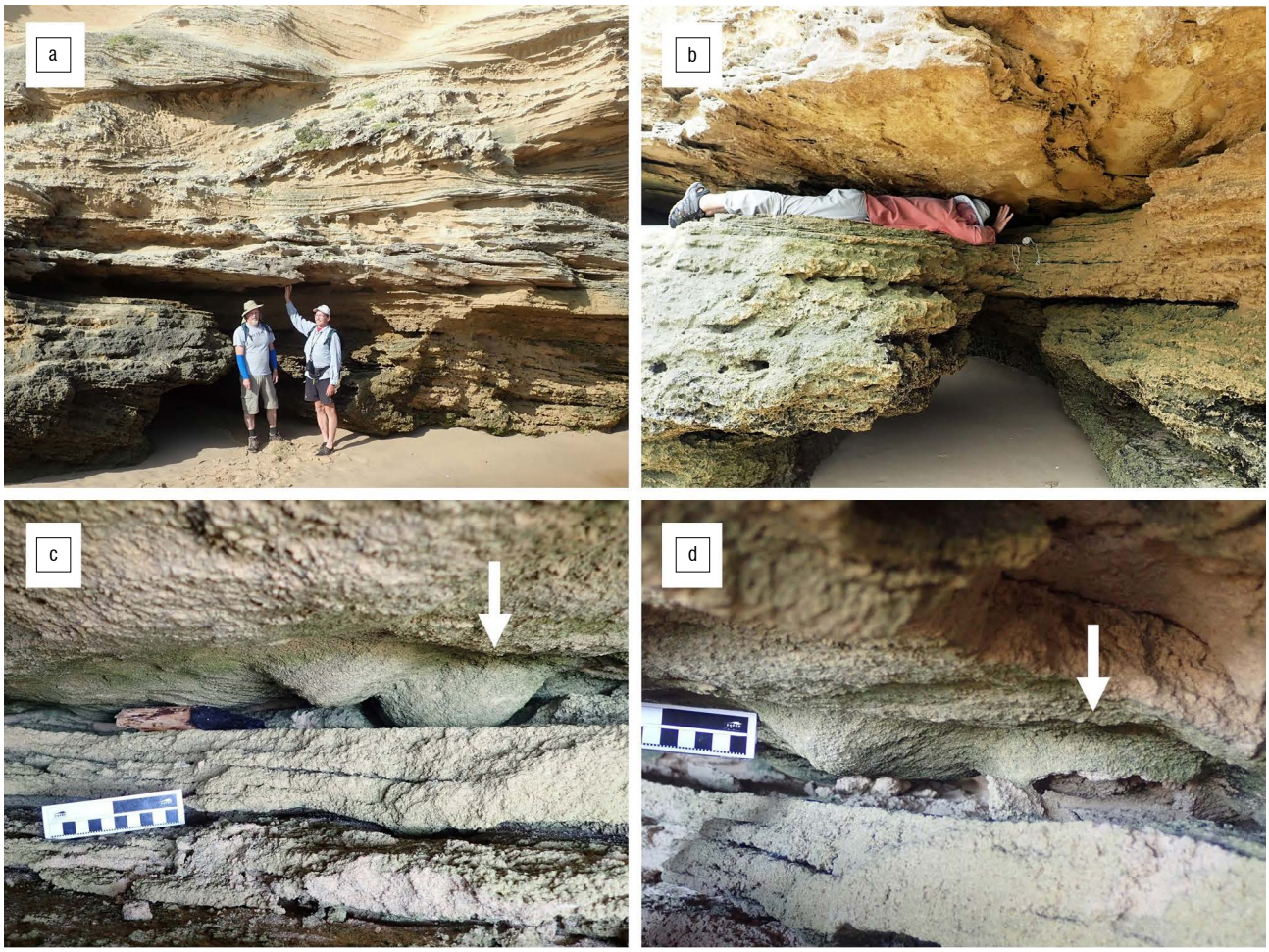


Figure 3: (a) Two figures at the entrance to the small cave that contains Site 1. (b) Distance between floor and ceiling is as little as 75 cm in the deeper portion of the cave. (c and d) Comparison views of Track A (indicated by arrows) before and after storm surge events – the track to the left of Track A is Track E; scale bars = 10 cm.

When viewed in sagittal section, the deepest points of these tracks are in their anterior (seaward) portions (Figure 4b). Photogrammetry was applied to the portion of the surface containing Track B and Track C (Figure 4c).

Although it is partly eroded, Track C appears to exhibit evidence of a medial longitudinal arch on its western aspect, and relative widening of its seaward (southern) portion along with vague evidence of a hallux orientated in the longitudinal access of the track. Consequently, it appears to be a natural cast of a left foot impression. We infer that the trackmaker was heading in an approximately southerly direction down a gently inclined surface.

This inference is borne out by the appearance of Track B in coronal section, which, although it shows little other morphological detail, is steeper on its western aspect than its eastern aspect, suggesting that it is a natural cast of a right foot impression (Figure 4d). The pace length between Track B and Track C is 82 cm.

A similar coronal section profile is evident in Track A, suggesting that it too is a natural cast of a right foot impression. If this is the case, and Tracks A and B form part of the same trackway, then the stride length between them is 165 cm.

If this analysis is accurate, then unfortunately the area where a left track would have occurred between tracks A and B has been eroded away. However, the remnants of a number of other tracks are evident on the ceiling in this eroded area, and layers in the side wall of the cave in this area show significant deformation, indicating that a substantial number of tracks must once have been present (Figure 4e).

An alternative interpretation is that one of these large eroded tracks, e.g. Track D, formed part of the trackway that includes Track B and Track C (Track D appears to be of approximately similar size to Tracks A, B and C, but is too eroded to permit detailed measurement). In this case, stride length would be ~145 cm.

More tracks are evident in hyporelief in the deeper, less accessible section of the cave, in the same axis as the putative trackway. However, these tracks do not appear to be well defined, and do not exhibit detailed morphology.

Of the tracks that are amenable to measurement, some appear to be significantly smaller than others, although Tracks A, B and C are similar in size, with characteristic hominin dimensions. For example, just west of Track A, and with similar orientation, is a probable track with a length of 13 cm (Track E). Close to the eroded area described above, one very eroded track (not shown in Figure 2) measures 15 cm in length and 8 cm in width, with a distinct vestigial outline and evidence of a medial longitudinal arch. The smallest natural cast feature, 9.5 cm in length and 5.5 cm in width, lies west of Track B. Because it contained no further distinguishing features, and although we assigned it a label (Track F), we could not be certain that it represented a hominin track.

In a seaward direction from Track C, and ~90 cm from it, another track is evident in hyporelief (Track G). Its orientation is almost perpendicular to that of the other tracks, with a bearing of 131°. It measures 24 cm in length and 10 cm in width (again consistent with the approximate dimensions of Tracks A, B and C), with a maximum depth of 5 cm and a minimum depth of 2.5 cm (Figure 5a). A heel-drag impression of 6 cm is apparent. As in the case of Track C, it appears broader in its anterior

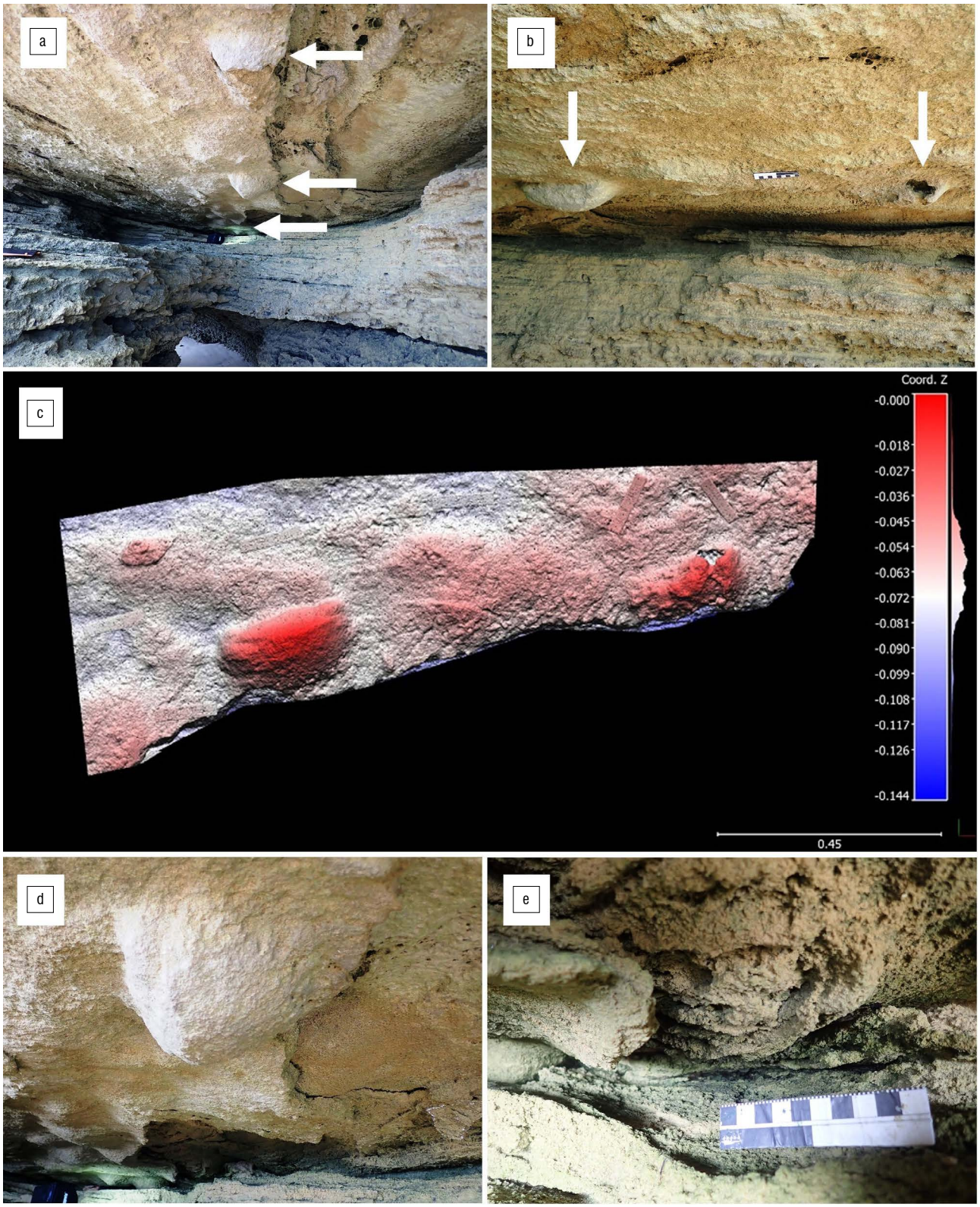


Figure 4: (a) Tracks in putative trackway are indicated with arrows, demonstrating narrow straddle: the top arrow points to Track C, the middle arrow points to Track B, and the bottom arrow points to Track A. (b) Tracks B (left) and C (right) are indicated with arrows; Track B is deeper in its anterior (seaward) portion; scale bar = 10 cm. (c) Photogrammetry colour mesh of Tracks B and C at Site 1, using 131 images. Track F is also visible, at top left. Photos were taken an average of 46.3 cm from the surface. The reprojection error is 0.488 pix. Vertical and horizontal scales are in metres. (d) Track B is steeper on its western (left in this photograph) aspect than its eastern aspect. (e) Track D; scale bar = 10 cm.

(distal) portion. Anterior to this, a vague hallux outline is evident in the longitudinal axis of the track, and a small ridge is evident posterior to the digit line. In this sense the track is better preserved than those with transmitted track features. For example, this track cast has rather steep walls, and so differs from the more rounded shapes of the transmitted tracks which therefore by definition show extra-morphological characteristics, and which we infer may have existed below Track G as transmitted features before they were eroded away. Additional tracks are evident, mostly in cross-section, at the junction of the ceiling with the eastern side wall of the cave, but these are not amenable to further interpretation. Photogrammetry was applied to Track G (Figure 5b).

Site 2: Goukamma Nature Reserve

Site 2 lies on the coastline of the Goukamma Nature Reserve, between the towns of Sedgefield and Buffels Bay. It was initially identified in 2012, but not analysed in detail until 2019. No significant deterioration appears to have taken place during this time interval. The track-bearing surface is evident in situ as hyporelief casts in cliffs that rise above the upper end of a beach, and as corresponding epirelief impressions on the upper surface of a fallen block at the foot of these cliffs (Figure 6a, 6b). The in-situ hyporelief surface has a strike measurement of 71°, with a dip measurement of 35°. The surface represents the infill layer of a Pleistocene dune, at the angle of repose of windborne sand. It measures 320 cm in width (east–west) and 260 cm from north to south. Its lowest point (at the northern end) is 270 cm above the basal sand level. Its highest point (at the southern end) is 418 cm above this sand level.

In contrast to Site 1, through the existence of both natural impressions (on the fallen block) and corresponding natural casts (on the in-situ overhang), we are encouraged to infer that the separation layer was probably the original track-bearing surface, and thus a temporarily exposed palaeosurface. Moreover, the separation resulted from the falling of the lower block, and perhaps a weakness at the palaeosurface, and not directly as the result of layers below the casts being actively eroded and dissolved away by forceful and direct marine erosion. This distinction is important, as it suggests the quality of preservation, although affected by some post-separation exposure, likely reflects the original quality of preservation on the track-bearing surface, and characteristics that the palaeosurface developed when exposed. These characteristics may have included subtle weathering of the tracks before burial.

The surface can be informally divided into four track-bearing areas. We created a trackway map of the hyporelief surface, which indicates these areas (Figure 7). To the west are 18 tracks which we interpret as having been registered in a north-to-south (upslope) direction. We label this Section A, and number these tracks from A1 to A18. It is evident that there are too many tracks in this section to have been formed by one trackmaker. Beginning near the northern (lower) area of these tracks, two approximately parallel trackways, containing six and five tracks, extend diagonally upslope in a southeasterly direction. The more southerly (more upslope) trackway contains six tracks. We label this Section B, and number these tracks from B1 to B6. The more northerly (more downslope) trackway contains five tracks. We label this Section C, and number these tracks from C1 to C5. Two further tracks, which we label D1 and D2, are evident further to the east in cross-section, where the track-bearing surface meets the cliff.

At a sectional distance of 80 cm below the main track-bearing surface there is another track-bearing layer, with unidentifiable tracks evident in cross-section and in hyporelief. A further probable track-bearing layer is evident in cross-section 70 cm below this layer.

The epirelief surface on the fallen block measures 270 cm from east to west and 280 cm from north to south. It is sometimes covered by sand. Not as many tracks are registered in epirelief, and their state of preservation is generally of inferior quality to those seen in hyporelief. However, this surface enables easier appreciation of numerous downslope displacement rims, which correspond to depressions on the hyporelief surface. It contains one track that is not apparent on the hyporelief surface, as the infill of this track must have broken off from the hyporelief surface when the block broke off. It is sharply outlined, and fills the depression on the epirelief surface (Figure 6c). Its anterior portion is wider than the posterior portion. In total there are therefore 32 tracks.

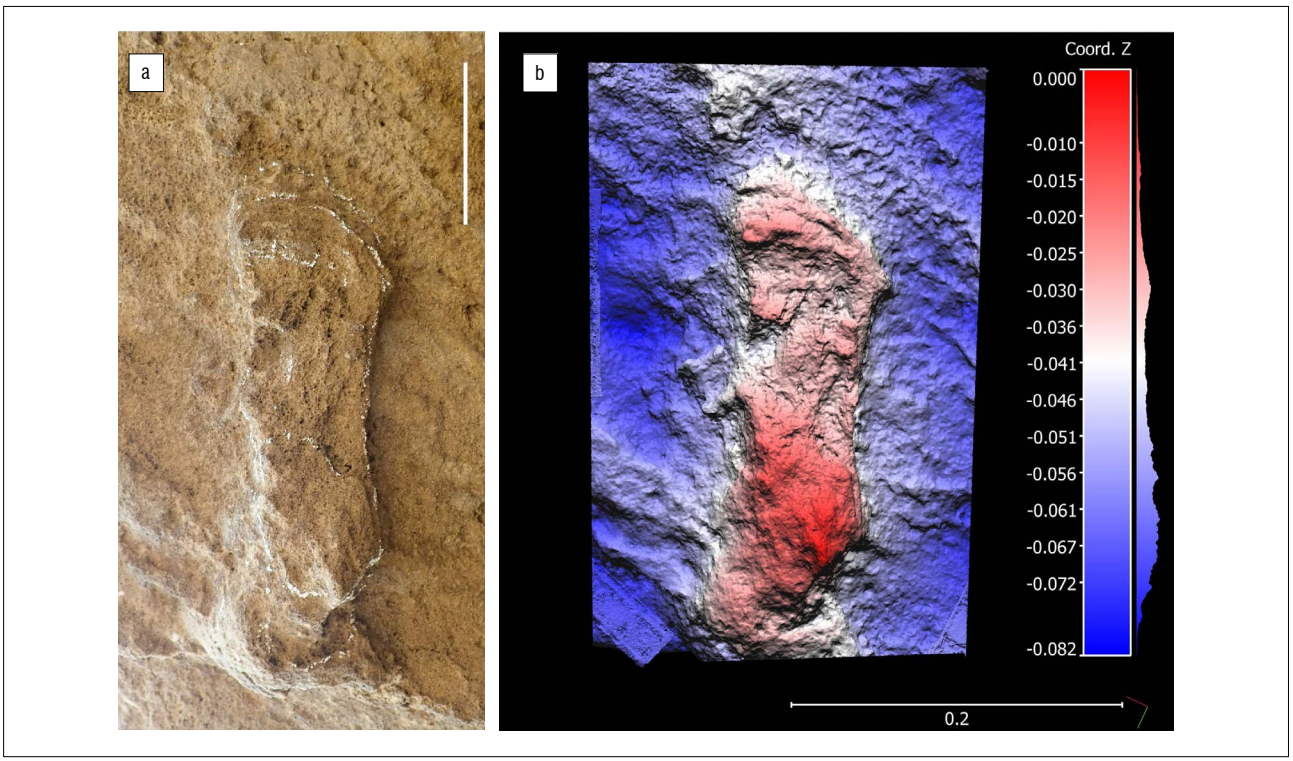


Figure 5: (a) Track G, lightly outlined in white chalk. White horizontal scale bar = 10 cm. (b) Photogrammetry colour mesh of Track G, using 59 images. Photos were taken an average of 34.2 cm from the surface. The reprojection error is 0.554 pix. Vertical and horizontal scales are in metres.

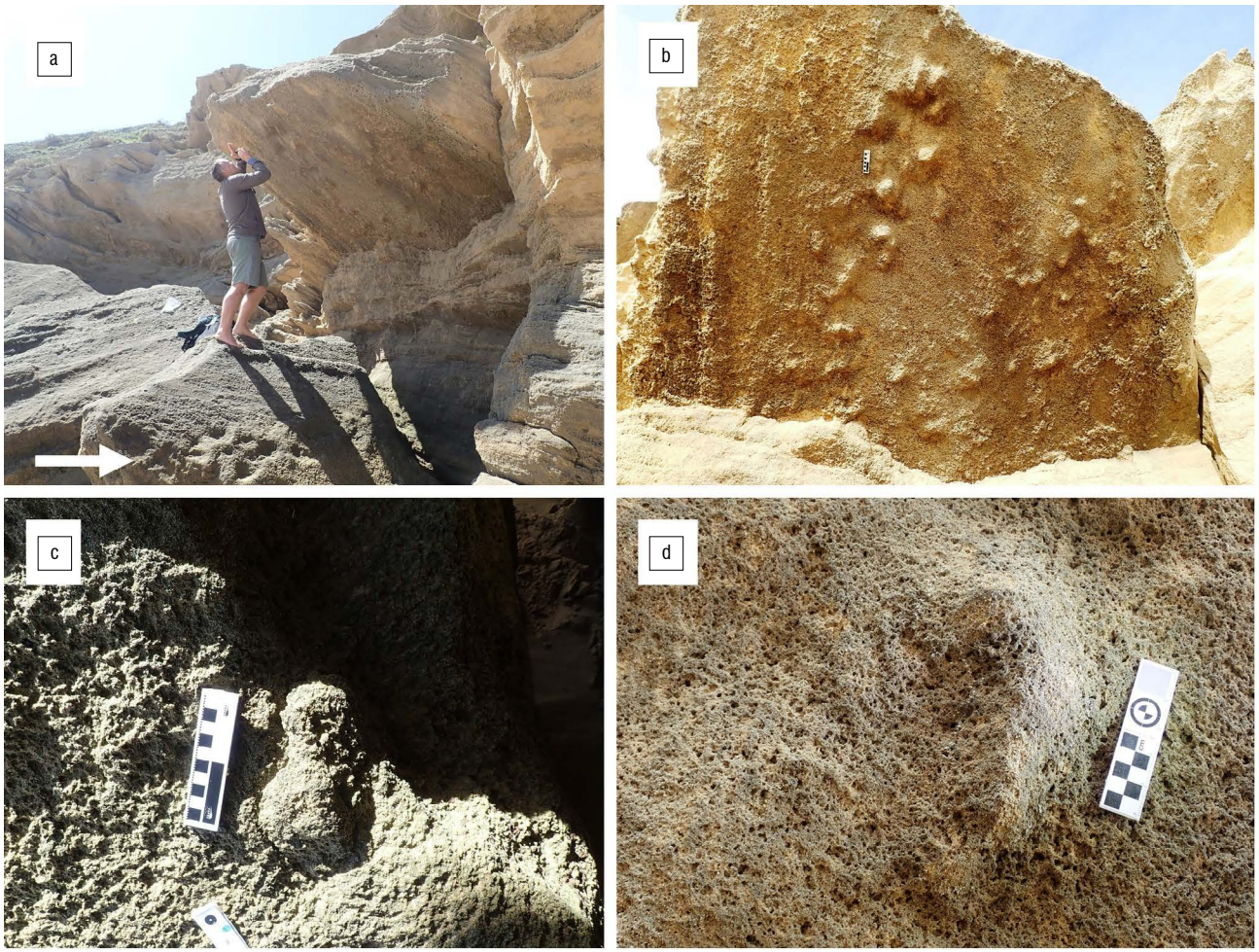


Figure 6: (a) The figure is standing on the top of the loose slab containing the tracks in epirelief (indicated by arrow) and analysing the tracks on the hyporelief surface. (b) The hyporelief surface of Site 2, viewed from a distance. (c) Close-up view of the track that is only evident on the epirelief surface. (d) Close-up view of Track A7.

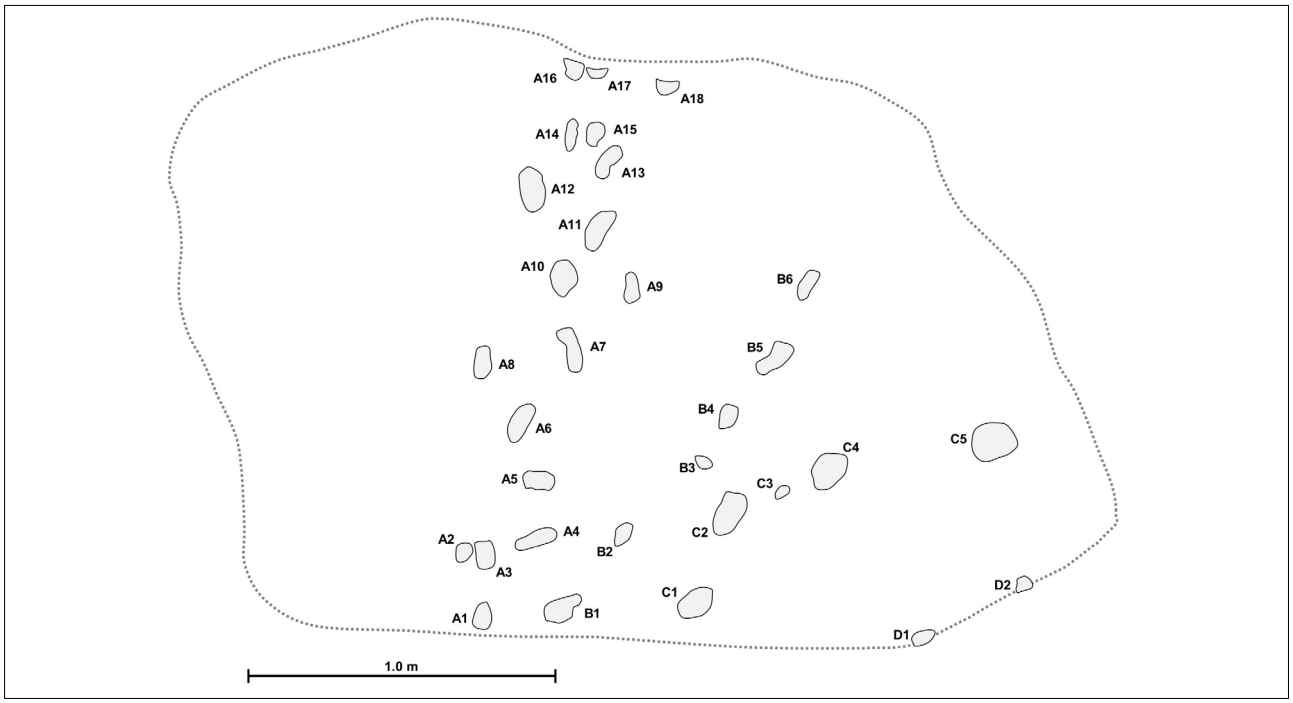


Figure 7: Trackway map of Site 2.

While some tracks in Section A are substantially longer than they are wide, and there is a size difference between some of the better-preserved tracks (suggesting the presence of more than one trackmaker), the outlines of others are poorly preserved or are truncated by the edge of the surface. Therefore, not all of these tracks were amenable to measurement. In some cases, where a probable medial longitudinal arch was present, left or right tracks could be identified (Figure 6d). Better preserved track morphology (including overall track dimensions and the presence of a medial longitudinal arch in some tracks) was noted in Trackway B. Measured dimensions, along with comments, are provided in Supplementary table 1. We applied photogrammetry to the hyporelief surface (Figure 8a) along with detailed views (Figure 8b, 8c), and the epirelief surface (Supplementary figure 1).

Sections A and B each provided evidence of a trackway. In the case of Section A, where many tracks were noted, with a similar longitudinal axis, but of varying size, a number of plausible trackways were identified, but none were unequivocal. Intertrack distances (plausible putative pace lengths) are presented in Supplementary table 2.

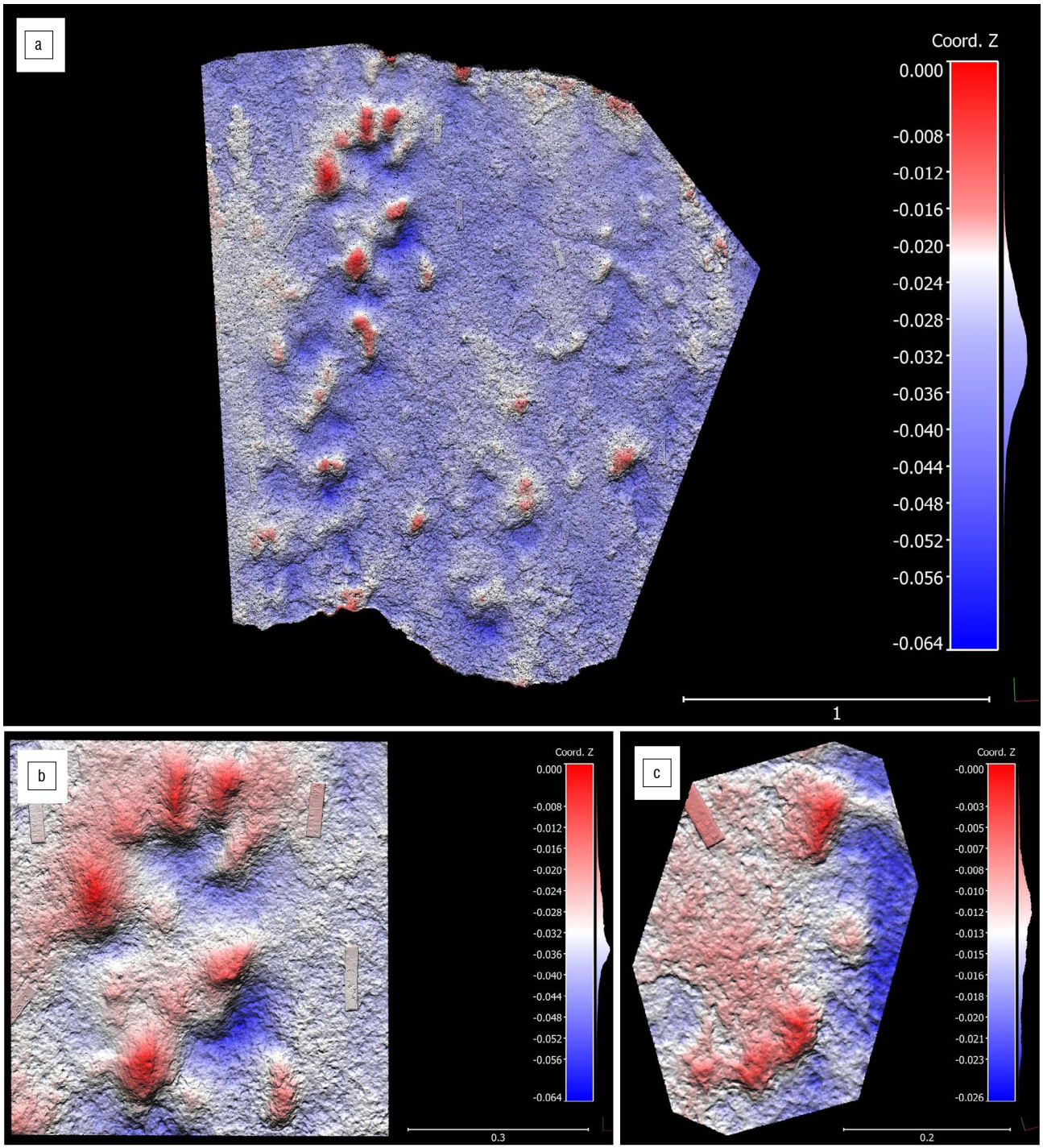


Figure 8: (a) Photogrammetry colour mesh of the Site 2 hyporelief surface, using 193 images. Photos were taken an average of 73.4 cm from the surface. The reprojection error is 0.427 pix. Vertical and horizontal scales are in metres. (b) Tight view of Tracks A9 – A15, with same photogrammetry metrics as (a). Vertical and horizontal scales are in metres. (c) Tight view of Track B5 and B6, showing probable left-right sequence, with same photogrammetry metrics as (a). Vertical and horizontal scales are in metres.

Site 3: Goukamma Nature Reserve

Site 3 lies on a southeast-facing truncation surface, with a dip of 25°, on the coastline of the Goukamma Nature Reserve, between the towns of Sedgefield and Buffels Bay. It was one of the sites described by Helm et al.[35], because of the presence of a number of sub-parallel grooves (one of which showed possible evidence of deliberate lengthening) as well as a number of smaller circular depressions, all clustered around an impression of what resembled the anterior portion of a left human foot (Figure 9a). It was postulated that this was an ammoglyph, possibly indicating messaging or foraging behaviour. The putative partial foot impression showed evidence of a hallux (aligned in the longitudinal axis) and two adjacent rows of digit impressions, explained by a slight pivot to the left (Figure 9b). Because only a single possible footprint was apparent, Helm et al.[35] were cautious not to propose any firm conclusion as to its origin.

Subsequently, further observations have been made at this site, following substantial scouring by wave action that led to the removal of a large amount of sand, which exposed underlying strata (Figure 9c). Two more partial impressions with patterns consistent with those of the anterior portions of human feet, including hallux and digits, have been identified. Unfortunately, both are truncated by surface margins. A 'stride length' of ~83 cm was measured between the newly identified (left) track at the bottom end of the main surface and the initially identified left partial foot impression. This left track was 9 cm wide, and was truncated 6 cm behind the front of the hallux, which measured ~2 cm in width. A ridge was present between the impressions of the hallux and putative digit II (Figure 9d).

A further putative partial hominin track was noted at the edge of a bedding plane surface a few centimetres lower than the main surface. It exhibited five impressions in epirelief (consistent with digit impressions of a right foot), separated by ridges, with a total width of 10 cm, suggesting that the trackmaker was heading upslope and slightly to the left of the fall line. The track was truncated 8 cm behind the front of the hallux.

Furthermore, grooves similar to those on the main surface were noted on these underlying surfaces. Two grooves, each ~12 cm long, and 10 cm apart at their upper ends and 7 cm apart at their lower ends, were noted on a surface situated 30 cm below the main surface. Maximum width of the left groove was 2 cm, and maximum width of the right groove was 3 cm. Faint rims were present beside these grooves. An infilled probable elephant pes impression, 37 cm long and 26 cm wide, was noted on this surface, at a distance of 200 cm from the grooves.

A further single deep groove was noted 220 cm to the right of these two grooves, on the same bedding plane surface. It was 10 cm in length, but truncated at its lower end. Maximum width was 2 cm, and a faint rim was present.

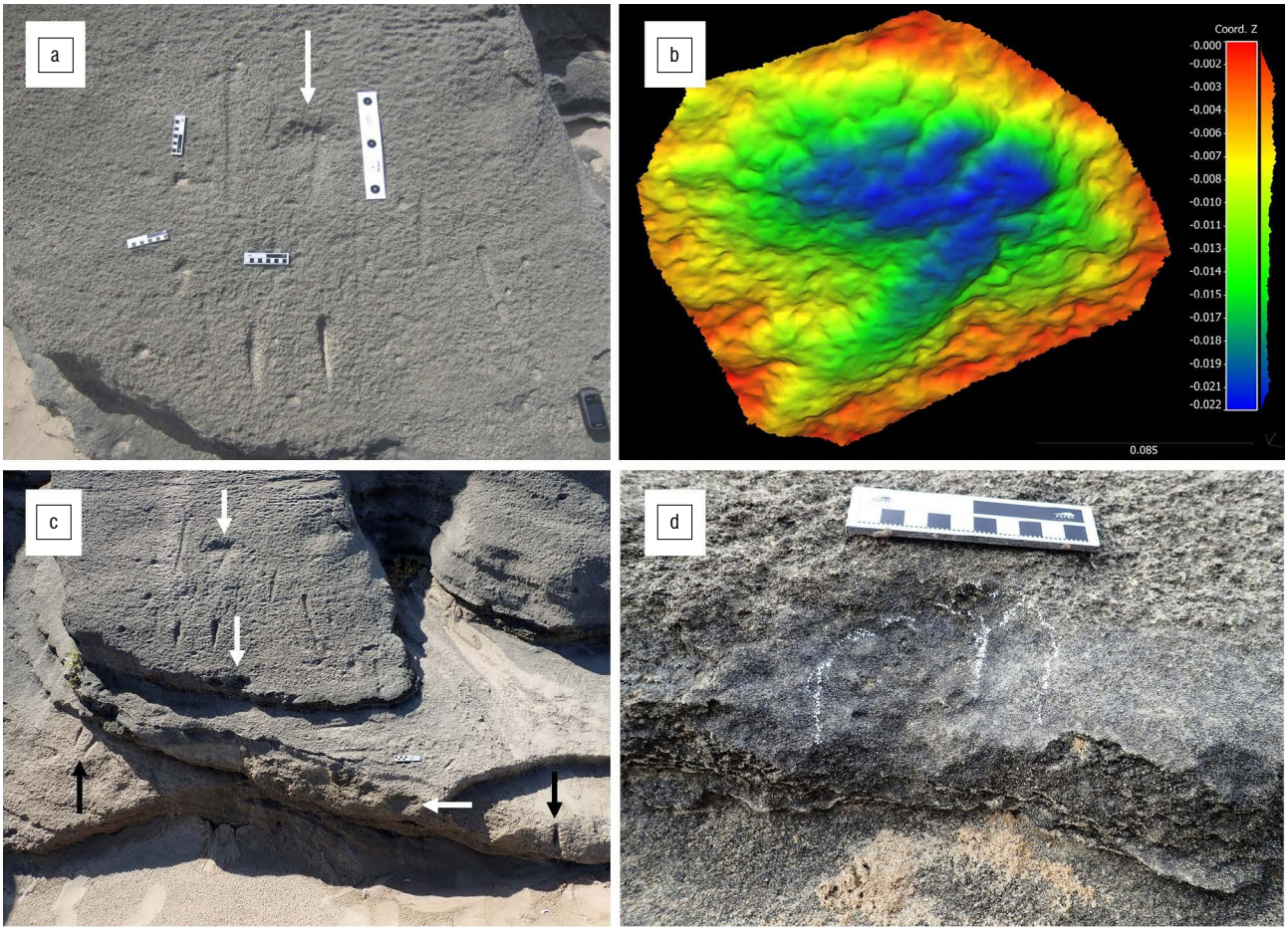


Figure 9: (a) The originally documented surface of Site 3, with central putative forefoot impression (indicated by arrow) and surrounding groove features; scale bars = 10 cm and 25 cm. (b) Photogrammetry colour mesh: tight tilt view of putative footprint on main surface of Site 3, using 27 images. Photos were taken an average of 0.466 m from the surface. The reprojection error is 0.243 pix. Vertical and horizontal scales are in metres. (c) Site 3 after extensive scouring by wave action; white arrows indicate the three probable partial human tracks; black arrows indicate newly exposed groove features; scale bar = 10 cm. (d) Newly identified probable partial human track, lightly outlined in white chalk; scale bar = 10 cm; this track is indicated by the middle white arrow in (c).

Site 4: Brenton-on-Sea

Site 4 is situated on the coast below the town of Brenton-on-Sea. A possible hominin tracksite is situated 330 m east of the previously documented hominin tracksite.[16] Tracks occur in a cliff-parallel, shore-parallel orientation, and are preserved in a cliff exposure, in hyporelief and cross-section, on a truncation surface that dips gently to the west. They exhibit more relief in their downslope (western) portions. Of the seven tracks that are evident, the western tracks occur in low relief (Figure 10a), the eastern tracks are deeper (Figure 10b), and the dimensions of others are not measurable. Considerable deformation of underlying layers is evident in cross-section below the eastern tracks.

An apparent narrow straddle was observed. Protrusions of the cliffs obscure the probable locations of some tracks. Distances measured between tracks might therefore represent stride lengths or multiples of pace lengths. We measured these, from east to west, as 191 cm, 66 cm, 110 cm, 73 cm, 63 cm and 70 cm. In only three cases could reliable track dimensions be obtained, and in only one case could track width be determined. From west to east:

- Length 23 cm, maximum depth 6 cm
- Length 21 cm, width 11-12 cm, maximum depth 6 cm
- Length 18 cm, maximum depth 5 cm

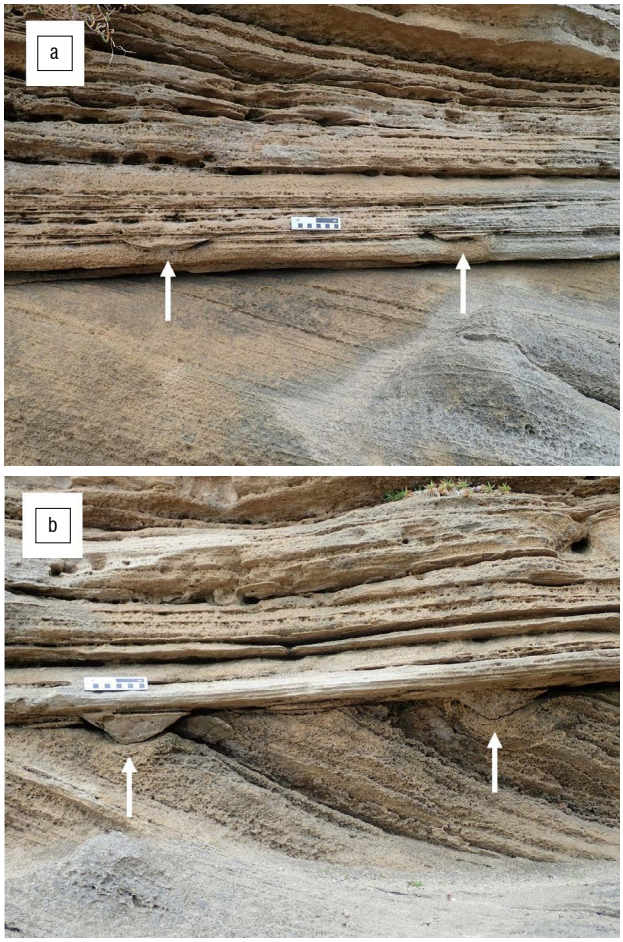


Figure 10: (a) The western tracks at Site 4, indicated by arrows, are in lower relief; scale bar = 10 cm. (b) The eastern tracks at Site 4, indicated by arrows, are deeper; scale bar = 10 cm.

Discussion

Tracksite interpretation

Sites 1 and 2 are similar in some respects to the previously described Brenton-on-Sea site.[16] In each case, multiple tracks (respectively 18, 32, and 40) are present, preserved in hyporelief. However, details of foot morphology, in particular with respect to digit morphology, are not well preserved at Sites 1 and 2. In all three cases, footprints of various sizes are present, indicating the presence of more than one trackmaker, and raising the possibility of family groups, as has been inferred at other hominin tracksites.[37,60,61] However, size variability is by no means random. On the contrary, at Site 1, Site 2 and the previously described Brenton-on-Sea site[16], there is considerable consistency in the size and shape of the larger tracks, which are characteristic of adult unshod humans. The smaller tracks indicate active juveniles (not infants) whose stature can be estimated, as discussed below. Putative pace lengths are greater at Site 1 than at Site 2. While this may in part be related to trackmaker size, it is probably also a reflection of downslope travel (Site 1) compared to upslope travel (Site 2).

Site 1 is also similar to the Brenton-on-Sea site[16] in that the track-bearing surface is susceptible to erosion through spring high tides and storm surges. In both cases, the 3D photogrammetric studies that have been performed have created a digital preservation archive, as these sites are not amenable to physical recovery, and replication through traditional means runs the risk of damaging the tracks.

Site 3 is in a different category: whereas only two or three footprint impressions are preserved (in epirelief), this is an advance over the single possible human forefoot impression reported at this site by Helm et al.[35] The presence of hallux and digit impressions consistent with a human trackmaker is significant, even though no complete track impressions are evident. Furthermore, the unexpected finding of grooves on multiple surfaces buttresses the contention that Site 3 is an ammoglyph site, suggesting an area that was repeatedly used by humans over an interval of time.

Site 4 has potential as a possible hominin tracksite, but nothing more conclusive can be stated until further features are exposed, through either an excavation or natural forces of erosion. It illustrates the challenges associated with aeolianite sites: cross-sectional exposures may indicate the presence of tracks, and may suggest a hominin trackmaker, but provide only limited information in the absence of larger surface exposures. In fact, all four sites described here are potentially amenable to further investigation through excavation, and the advantages and disadvantages (including the risk to personal safety in the case of caves and overhangs) of such an approach, compared with monitoring of the effects of erosive processes, need to be weighed up.

Samples have been taken for OSL dating from Site 1, Site 2 and Site 3, as well as from the previously described Brenton-on-Sea site.[16] Based upon results of previous dating studies from the Wilderness Embayment, we anticipate that the results obtained from these samples will likely be from sub-stages of MIS 5.

The role of substrate

Tuttle's criteria[33] apply, in our view, to near-ideal substrates and track preservation. While meeting them might be a laudable goal, strict adherence to them will lead to under-identification of hominin tracksites. Site 1 and Site 2 illustrate the challenges of identifying hominin tracks that were registered in unconsolidated surfaces of sand.

In the case of unconsolidated sand substrates, there is an ideal for track preservation between the extremes of too firm and too soft, and too wet and too dry. Moist dune sand more readily preserves identifiable footprints.[62,63] Slope steepness plays an additional role in footprint morphology and in the presence, position and appearance of displacement rims.[16] Bennett and Morse[18] summarised the role of substrate differences with respect to hominin track morphology.

Essentially, if moisture content is too high or too low, or if substrates are too firm or too soft, track morphology will be sub-optimally preserved.

The relative lack of track detail evident at Sites 1 and 2 is most likely a reflection of the tracks (some of which we interpret as transmitted tracks) having been made in fairly soft dune surfaces, compounded by some degree of post-exhumation weathering and erosion. The substantial downslope displacement rims at Site 2 are consistent with the steepness of the dune slope. These sites may be compared with the example of the Happisburgh tracksite in the United Kingdom[64], where 50 tracks with a hominin outline were described, but only two exhibited evidence of digit traces.

Distinguishing features

At Site 1 and Site 2 we used a combination of the following features in our identification, in addition to searching for standard features such as medial longitudinal arch, and hallux and other digits:

- straight trackway
- narrow straddle
- long pace length
- tracks are much longer than they are wide
- tracks are broader anteriorly (forefoot) than posteriorly (hindfoot)
- coronal section profile is steeper laterally than medially.

While each of these in isolation may not be diagnostic, in combination we contend that they may allow the identification of hominin tracks, even in softer sub-optimal substrates.

Knowledge of other candidate Pleistocene trackmakers from the region, that could conceivably create tracks with these features, is imperative. In this regard, other hominins, such as *Homo naledi or Homo helmei*, cannot be excluded, but are unlikely, and there are no comparisons to refer to between putative tracks of these species and *H. sapiens.*[16,21] The track dimensions are much too large to have been made by other extant primates in the region – *Papio ursinus* (chacma baboons) or *Chlorocebus pygerythrus* (vervet monkeys) – which are in any event quadrupedal.[65]

Overstepping by large vertebrate trackmakers may lead to compound manus-pes track sets that are significantly longer than they are wide. Equids employ such a gait pattern more frequently than bovids.[65] In this gait pattern, the hind feet, rather than being placed in the impressions made by the front feet (direct register), are placed just ahead of them; thus the pes 'oversteps' the manus. However, tracks of the hind feet of equids and large bovids are typically slightly narrower than the tracks of their front feet. In order for the anterior portion of such a compound track to appear slightly wider than the posterior portion (and thus resemble a human footprint), the pes would need to be placed behind the manus track (understepping), a phenomenon associated with a much slower gait speed.[65] Such a gait would be associated with a much shorter pace length. Knowledge of typical gait patterns of equids (e.g. trot and gallop patterns) is also useful: where such patterns are evident, a hominin trackmaker can be excluded. We applied these concepts at Site 1 and Site 2, and did not find evidence to support equids or bovids as trackmakers.

The larger the surface exposure, and the greater the number of tracks and trackways, the less likelihood there is of identification error. Duveau et al.[25] analysed a tracksite containing 257 footprints in France, attributed to *Homo neanderthalensis*. Enough morphological detail was present to permit morphometric analysis, and to lead to the conclusion that the tracks represented a single brief occupation event and were made by a small group, with a majority of children. Thus far, a site with such a large number of tracks and such a level of morphological detail has not been encountered on the Cape south coast. The most likely sites for substantial bedding plane exposures that might provide this number of tracks and this level of detail may be on cave ceilings, as described from Robberg.[66]

Of the four sites described here, one occurs in epirelief, two occur predominantly in hyporelief, and one occurs in both epirelief and hyporelief. The previously noted trend for the southern African sites to be predominantly preserved in hyporelief thus continues. Site 2, where tracks in epirelief occur on the upper surface of a large fallen slab, and the infill of the original surface is preserved above, in situ, in hyporelief, is an unusual and possibly globally unique phenomenon for hominin tracksites.

Trackmaker stature

Bennett and Morse[18] described the challenges and pitfalls inherent in estimating stature, velocity and mass from hominin trackway parameters, and the potential that results for over-interpretation. Nonetheless, stature estimates have been provided for the previously described southern African hominin tracksites: Roberts[17] used a formula derived from global mean data used by Mietto et al.[67], whereby trackmaker height = footprint length x 6.67. This led to height estimates of 128.06 cm for the Nahoon trackmaker and 152.07 cm for the Langebaan trackmaker.[17] This formula, applied to the hominin tracks at the Brenton-on-Sea site, yielded height estimates of 153.4 cm for the largest tracks and ~116 cm for the smaller tracks.

Applying this formula to the larger tracks at Site 1, using a track length of 24 cm, a height estimate of ~160 cm is obtained. For Site 2, the longest tracks (20 cm) yield a height estimate of ~133 cm, and the shorter tracks (13 cm) yield an estimate of ~87 cm. We acknowledge that the lack of fine preservation of the tracks contributes to the level of uncertainty associated with such estimates.

Romano et al.[37], in describing very well preserved hominin tracks in an Italian cave, used a foot length/stature ratio of 0.1541, calculated from a sample of Upper Palaeolithic adults. Using this ratio, estimated height = footprint length x 6.49, the above estimates need to be reduced by between 2% and 3%. For example, the height estimate for the trackmaker that created the larger tracks at Site 1 is ~156 cm rather than 160 cm, a difference of only 2.5%.

Evidence of social behaviour

At Site 1, Site 2 and the Brenton-on-Sea site[16], there is evidence of social behaviour; i.e. none of the sites represents the tracks of lone individuals. Also, given the relatively small size of the palaeosurfaces, the number of individuals represented is likely a minimum number, and there is evidence at these sites of several individuals moving in the same direction. Furthermore, there is no sedimentological evidence against the different trackways having been registered at the same time, i.e. the tracks appear similar in depth and quality of preservation. Together, these sites "fill out" the palaeo-geographical distribution range of hominin activity in the region, complementing the archaeological record of occupation sites along this coast.

Conclusion

If the Langebaan site is included, southern Africa now boasts six hominin footprint sites representing the tracks of unshod hominins inferred to represent *Homo sapiens*. These are the six oldest hominin footprint sites thus far reported that are attributed to our own species. They form an important part of what remains a sparse global hominin track record. Although track quality may not be optimal, these sites prove the capacity of coastal aeolianites to reliably preserve such features, and to complement the Middle Stone Age hominin record obtained from other avenues of study. When we survey the rich ichnological record of this region, the conclusion is inescapable: hominin tracks are more common than previously supposed.

Acknowledgements

We appreciate the support of Jack Carrigan, Linda Helm, Nigel Mathews, the staff of Garden Route National Park, and the staff of Goukamma Nature Reserve.

Competing interests

We declare that there are no competing interests.

Authors' contributions

C.W.H.: Lead author, corresponding author, conceptualisation, data collection, data analysis, project leadership, track analysis, photogrammetry. M.G.L.: Second author, site analysis, conceptualisation, data analysis, track analysis, review of drafts and revisions. H.C.C.: Conceptualisation, sample analysis, site analysis, data analysis, contribution on geological context, field stratigraphy, review of drafts and revisions. J.C.D.V.: Site visit, data analysis, track analysis, review of drafts and revisions. M.G.D.: Site analysis, track analysis, photogrammetry, review of drafts and revisions. C.J.Z.H.: Site analysis, data analysis, photogrammetry, review of drafts and revisions. G.H.H.T.: Site analysis, track analysis, site mapping, review of drafts and revisions.

Data availability

The data supporting the results of this study are available upon request to the corresponding author.

References

1. Henshilwood CS, d'Errico F, Van Niekerk KL, Coquinot Y, Jacobs Z, Lauritzen S-E, et al. A 100,000-year-old ochre-processing workshop at Blombos Cave, South Africa. Science. 2011;334:219–222. https://.org/10.1126/science.1211535
2. Marean CW. Pinnacle Point Cave 13B (Western Cape Province, South Africa) in context: The Cape Floral Kingdom, shellfish, and modern human origins. J Hum Evol. 2010;59:425–443. https://.org/10.1016/j.jhevol.2010.07.011
3. Wurz S. Modern behaviour at Klasies River. South African Archaeological Society Goodwin Series. 2008;10:150–156.
4. Marean CW, Bar-Matthews M, Bernatchez J, Fisher E, Goldberg P, Herries AIR, et al. Early human use of marine resources and pigment in South Africa during the Middle Pleistocene. Nature. 2007;449:905–908. https://doi.org/10.1038/nature06204
5. d'Errico F, Henshilwood C, Vanhaeren M, Van Niekerk, K. *Nassarius kraussianus* shell beads from Blombos Cave: Evidence for symbolic behaviour in the Middle Stone Age. J Hum Evol. 2005;48:3–24. https://doi.org/10.1016/j.jhevol.2004.09.002
6. Brown KS, Marean CW, Jacobs Z, Schoville BJ, Oestmo S, Fisher EC, et al. An early and enduring advanced technology originating 71 000 years ago in South Africa. Nature. 2012;491:590–593. https://doi.org/10.1038/nature11660
7. Brown KS, Marean CW, Herries AIR, Jacobs Z, Tribolo C, Braun D, et al. Fire as an engineering tool of early modern humans. Science. 2009;325:859–862. http://dx.doi.org/10.1126/science.1175028
8. Henshilwood C, d'Errico F, Marean C, Milo R, Yates R. An early bone tool industry from the Middle Stone Age at Blombos Cave, South Africa: Implications for the origins of modern human behaviour, symbolism and language. J Hum Evol. 2001;41:632–678. https://doi.org/10.1006/jhev.2001.0515
9. Henshilwood CS, d'Errico F, Yates R, Jacobs Z, Tribolo C, Duller GAT, et al. Emergence of modern human behavior: Middle Stone Age engravings from South Africa. Science. 2002;295:1278–1280. https://doi.org/10.1126/science.1067575
10. Watts I. The pigments from Pinnacle Point Cave 13B, Western Cape, South Africa. J Hum Evol. 2010;59(3–4):392–411. https://doi.org/10.1016/j.jhevol.2010.07.006
11. Henshilwood CS, d'Errico F, Van Niekerk KL, Dayet L, Queffelec A, Pollarolo L. An abstract drawing from the 73,000-year-old levels at Blombos Cave, South Africa. Nature. 2018;562:115–118. https://doi.org/10.1038/s41586-018-0514-3
12. Helm CW, Cawthra HC, Cowling RM, De Vynck JC, Lockley MG, Marean CW, et al. Pleistocene vertebrate tracksites on the Cape south coast of South Africa and their potential palaeoecological implications. Quat Sci Rev. 2019;235:105857. https://doi.org/10.1016/j.quascirev.2019.07.039
13. Helm CW, Cawthra HC, Cowling RM, De Vynck JC, Marean CW, McCrea RT, et al. Palaeoecology of giraffe tracks in Late Pleistocene aeolianites on the Cape south coast. S Afr J Sci. 2018;114(1/2), Art. #2017-0266, 8 pages. http://dx.doi.org/10.17159/sajs.2018/20170266
14. Helm CW, Cawthra HC, Combrink X, Helm CJZ, Rust R, Steer W, et al. Pleistocene large reptile tracks and probable swim traces on South Africa's Cape south coast. S Afr J Sci. 2020;116(3/4), Art. #6542, 8 pages. https://doi.org/10.17159/sajs.2020/6542
15. Lockley MG, Cawthra HC, De Vynck JC, Helm CW, McCrea RT, Nel R. New fossil sea turtle trackway morphotypes from the Pleistocene of South Africa highlight role of ichnology in turtle palaeobiology. Quat Res. 2019;92(3):626–640. https://doi.org/10.1017/qua.2019.40
16. Helm CW, McCrea RT, Cawthra HC, Cowling RM, Lockley MG, Marean CW, et al. A new Pleistocene hominin tracksite from the Cape south coast, South Africa. Sci Rep. 2018;8, Art. #3772, 13 pages. http://dx.doi.org/10.1038/s41598-018-22059-5
17. Roberts DL. Last Interglacial hominid and associated vertebrate fossil trackways in coastal eolianites, South Africa. Ichnos. 2008;15(3):190–207. http://dx.doi.org/10.1080/10420940802470482
18. Bennett MR, Morse SA. Human footprints: Fossilised locomotion? Springer; 2014. https://doi.org/10.1007/978-3-319-08572-2
19. Lockley M, Roberts G, Kim JY. In the footprints of our ancestors: An overview of the hominid track record. Ichnos. 2008;15(3–4):106–125. http://dx.doi.org/10.1080/10420940802467835
20. Lockley M, Meldrum J, Kim JY. Major events in hominin evolution. In: Mángano MG, Buatois LA, editors. The trace-fossil record of major evolutionary events. Volume 2" Mesozoic and Cenozoic. Topics in Geobiology 40. Dordrecht: Springer; 2016. p. 411–418.
21. Helm CW, Lockley MG, Cole K, Noakes TD, McCrea RT. Hominin tracks in southern Africa: A review and an approach to identification. Palaeont Afr. 2019;53:81–96.
22. Lockley MG, Kim JY, Roberts G. The Ichnos project: A re-evaluation of the hominid track record. In: Lucas SG, Spielmann JA, Lockley MG, editors. Cenozoic vertebrate tracks and traces. New Mexico Museum of Natural History & Sciences Bulletin. 2007;42:79–90.
23. Kim JY, Kim KS, Lockley MG, Matthews N. Hominid ichnotaxonomy: An exploration of a neglected discipline. Ichnos 2008;15:126–139. https://doi.org/10.1080/10420940802467868
24. Gierliński GD, Niedźwiedzki G, Lockley MG, Athanassiou A, Fassoulas C, Dubicka Z, et al. Possible hominin footprints from the late Miocene (c. 5.7 Ma) of Crete? Proc Geol Assoc. 2017;128(5–6):697–710. https://doi.org/10.1016/j.pgeola.2017.07.006
25. Duveau J, Berillon G, Verna C, Laisné G, Cliquet D. The composition of a Neandertal social group revealed by the hominin footprints at Le Rozel (Normandy, France). Proc Natl Acad Sci USA. 2019;116(39):19409–19414. https://doi.org/10.1073/pnas.1901789116
26. Belvedere M, Farlow JO. A numerical scale for quantifying the quality of preservation of vertebrate tracks. In: Falkingham PL, Marty D, Richter A, editors. Dinosaur tracks: The next steps. Bloomington and Indianapolis, IN: Indiana University Press; 2016. p. 92–99.
27. Mountain ED. Footprints in calcareous sandstone of Nahoon Point. S Afr J Sci. 1966;62:103–111.
28. Le Roux FG. Lithostratigraphy of the Nahoon Formation (Algoa Group). South African Committee for Stratigraphy Lithostratigraphic Series 9. Pretoria: Department of Mineral and Energy Affairs; 1989.
29. Jacobs Z, Roberts DL. Last interglacial age for aeolian and marine deposits and the Nahoon fossil human footprints, southeast coast of South Africa. Quat Geochron. 2009;4:160–169. https://doi.org/10.1016/j.quageo.2008.09.002
30. Roberts D, Berger LR. Last Interglacial (c. 117 kyr) human footprints from South Africa. S Afr J Sci. 1997;93(8):349–350.
31. Roberts DL, Botha GA, Maud RR, Pether J. Coastal Cenozoic deposits. In: Johnson MR, Annhauser CR, Thomas RJ, editors. The geology of South Africa. Johannesburg: Geological Society of South Africa/Council for Geoscience; 2006. p. 605–628.

32. Malan JA. Lithostratigraphy of the Waenhuiskrans Formation (Bredasdorp Group). South African Committee for Stratigraphy Lithostratigraphic Series 8. Pretoria: Department of Mineral and Energy Affairs; 1989.

33. Tuttle RH. Footprint clues in hominid evolution and forensics: Lessons and limitations. Ichnos. 2008;15:158–165.

34. Morse SA, Bennett MR, Liutkus-Pierce C, Thackeray F, McClymont J, Savage R, et al. Holocene footprints in Namibia: The influence of substrate on footprint variability. Am J Phys Anthropol. 2013;151:265–279. https://doi.org/10.1002/ajpa.22276

35. Helm CW, Cawthra HC, De Vynck JC, Helm CJZ, Rust R, Stear W. Patterns in the sand: A Pleistocene hominin signature along the South African coastline? Proc Geol Assoc. 2019;130(6):719–740. https://doi.org/10.1016/j.pgeola.2019.08.004

36. Pastoors A, Lenssen-Erz T, Ciqae T, Kxunta U, Thao T, Bégouën R, et al. Tracking in caves: Experience based reading of Pleistocene human footprints in French caves. Camb Archaeol J. 2015;25:551–564. https://doi.org/10.1017/S0959774315000050

37. Romano M, Citton P, Salvador I, Arobba D, Rellini I, Firpo M, et al. A multidisciplinary approach to a unique Palaeolithic human ichnological record from Italy (Bàsura Cave). eLife. 2019;8, e45204. https://doi.org/10.7554/eLife.45204

38. Fairbridge RW, Johnson DL. Eolianite. In: Fairbridge RW, Bourgeois J, editors. The encyclopedia of sedimentology. Stroudsburg: Dowden, Hutchinson and Ross; 1978.

39. Malan JA. Lithostratigraphy of the Klein Brak Formation (Bredasdorp Group). South African Committee for Stratigraphy Lithostratigraphic Series 13. Pretoria: Department of Mineral and Energy Affairs; 1991.

40. Roberts DL, Bateman MD, Murray-Wallace CV, Carr AS, Holmes PJ. Last interglacial fossil elephant trackways dated by OSL/AAR in coastal aeolianites, Still Bay, South Africa. Palaeogeogr Palaeoclimatol Palaeoecol. 2008;257(3):261–279. http://dx.doi.org/10.1016/j.palaeo.2007.08.005

41. Roberts DL, Karkanas P, Jacobs Z, Marean CW, Roberts RG. Melting ice sheets 400,000 yr ago raised sea level by 13 m: past analogue for future trends. Earth Planet Sci Lett. 2012;357–358:226–237. https://doi.org/10.1016/j.epsl.2012.09.006

42. Carr AS, Bateman MD, Roberts DL, Murray-Wallace CV, Jacobs Z, Holmes PJ. The last interglacial sea-level high stand on the southern Cape coastline of South Africa. Quat. Res. 2010;73:351–363. https://doi.org/10.1016/j.yqres.2009.08.006

43. Carr AS, Bateman MD, Cawthra HC, Sealy J. First evidence for onshore Marine Isotope Stage 3 aeolianite formation on the southern Cape coastline of South Africa. Mar Geol. 2019;407:1–15. https://doi.org/10.1016/j.margeo.2018.10.003

44. Bateman MD, Carr AS, Dunajko AC, Holmes PJ, Roberts DL, Mclaren SJ, et al. The evolution of coastal barrier systems: A case study of the Middle-Late Pleistocene Wilderness barriers, South Africa. Quat Sci Rev. 2011;30:63–81. https://doi.org/10.1016/j.quascirev.2010.10.003

45. Cawthra HC, Jacobs Z, Compton JS, Fisher EC, Karkanas P, Marean CW. Depositional and sea-level history from MIS 6 (Termination II) to MIS 3 on the southern continental shelf of South Africa. Quat Sci Rev. 2018;181:156–172. https://doi.org/10.1016/j.quascirev.2017.12.002

46. Roberts DL, Cawthra HC, Musekiwa C. Dynamics of late Cenozoic aeolian deposition along the South African coast: A record of evolving climate and ecosystems. In: Martini IP, Wanless HR, editors. Sedimentary coastal zones from high to low latitudes: Similarities and differences. Special Publication of the Geological Society of London 388. London: Geological Society of London; 2013. p. 353–387 http://dx.doi.org/10.1144/SP388.11

47. Flügel E. Microfacies of carbonate rocks: analysis, interpretation and application. Berlin, Heidelberg, New York: Springer-Verlag; 2004.

48. Illenberger WK. The geomorphic evolution of the Wilderness dune cordons, South Africa. Quat Int. 1996;33:11–20. https://doi.org/10.1016/1040-6182(95)00099-2

49. Bateman MD, Holmes PJ, Carr AS, Horton BP, Jaiswal MK. Aeolianite and barrier dune construction spanning the last two glacial-interglacial cycles from the southern Cape coast, South Africa. Quat Sci Rev. 2004;23:1681–1698. https://doi.org/10.1016/j.quascirev.2004.02.001

50. Fisher EC, Bar-Matthews M, Jerardino A, Marean CW. Middle and Late Pleistocene paleoscape modeling along the southern coast of South Africa. Quat Sci Rev. 2010;29:1382–1398. http://dx.doi.org/10.1016/jquascirev.2010.01.015

51. Cawthra HC, Bateman MD, Carr AS, Compton JS, Holmes PJ. Understanding Late Quaternary change at the land-ocean interface: A synthesis of the evolution of the Wilderness coastline, South Africa. Quat Sci. Rev. 2014;99(1):210–223. https://doi.org/10.1016/j.quascirev.2014.06.029

52. Marean CW, Cawthra HC, Cowling RM, Esler KJ, Fisher E, Milewski A, et al. Stone Age people in a changing South African Greater Cape Floristic Region. In: Allsopp N, Colville JF, Verboom GA, editors. Fynbos: Ecology, evolution, and conservation of a megadiverse region. Oxford: Oxford University Press; 2014. p. 164–199. http://dx.doi.org/10.1093/acprof:oso/9780199679584.001.0001

53. Waelbroeck C, Frank N, Jouzel J, Parrenin F, Masson-Delmotte V, Genty D. Transferring radiometric dating of the last interglacial sea level highstand to marine and ice core records. Earth Planet Sci Lett. 2008;265:183–194. https://doi.org/10.1016/j.epsl.2007.10.006

54. Carr AS, Bateman MD, Holmes PJ. Developing a 150ka luminescence chronology for the barrier dunes of the southern Cape, South Africa. Quat Geochron. 2007;2(1):110–116. https://doi.org/10.1016/j.quageo.2006.09.002

55. Fleming K, Johnston P, Zwartz D, Yokoyama Y, Lambeck K, Chappel J. Refining the eustatic sea-level curve since the Last Glacial Maximum using far- and intermediate-field sites. Earth Planet Sci Lett. 1998;163(1):327–342. https://doi.org/10.1016/S0012-821X(98)00198-8

56. Newton AR, Shone RW, Booth PWK. The Cape Fold Belt. In: Johnson MR, Annhauser CR, Thomas RJ, editors. The geology of South Africa. Johannesburg: Geological Society of South Africa/Council for Geoscience; 2006. p. 521–531.

57. Leonardi G. Glossary and manual of tetrapod footprint paleoichnology. Brasilia: Ministry of Mines and Energy, Department of Mineral Production; 2017.

58. Falkingham PL, Bates KT, Avanzini M, Bennett M, Bordy EM, Breithaupt BH, et al. A standard protocol for documenting modern and fossil ichnological data. Palaeontology. 2018;61(4):469–480. https://doi.org/10.1111/pala.12373

59. Falkingham PL, Marty D, Richter A, editors. Dinosaur tracks: The next steps. Bloomington and Indianapolis, IN: Indiana University Press; 2016.

60. Citton P, Romano M, Salvador I, Avanzini M. Reviewing the upper Pleistocene human footprints from the 'Sala dei Misteri' in the Grotta della Bàsura (Toirano, northern Italy) cave: An integrated morphometric and morpho-classificatory approach. Quat Sci Rev. 2017;169:50–64. https://doi.org/10.1016/j.quascirev.2017.05.016

61. Pastoors A, Lenssen-Erz T, Ciqae T, Kxunta U, Thao T, BégouënR, et al. Tracking in caves: Experience based reading of Pleistocene human footprints in French caves. Camb Archaeol J. 2015;25:551–564. https://doi.org/10.1017/S0959774315000050

62. McKee ED. Tracks that go uphill. Plateau. 1944;16:61–72.

63. Lewis DW, Titheridge DG. Small scale sedimentary structures resulting from foot impressions in dune sand. J Sediment Petrol. 1978;48:835–838.

64. Ashton N, Lewis SG, De Groote I, Duffy SM, Bates M, Bates R, et al. Hominin footprints from Early Pleistocene deposits at Happisburgh, UK. PLoS ONE. 2014;9(2), e88329. https://doi.org/10.1371/journal.pone.0088329

65. Van den Heever A, Mhlongo R, Benadie K. Tracker manual – a practical guide to animal tracking in southern Africa. Cape Town: Struik Nature; 2017.

66. Helm CW, Cawthra HC, Hattingh R, Hattingh S, McCrea RT, Thesen GHH. Pleistocene trace fossils of Robberg Nature Reserve. Palaeont Afr. 2019;54:36–47. http://wiredspace.wits.ac.za/handle/10539/28633

67. Mietto P, Avanzini M, Rolandi G. Palaeontology: Human footprints in Pleistocene volcanic ash. Nature. 2003;422:133. https://doi.org/10.1038/422133a

AUTHOR:
Neil Overy[1]

AFFILIATION:
[1]Environmental Humanities South, University of Cape Town, Cape Town, South Africa

CORRESPONDENCE TO:
Neil Overy

EMAIL:
neilovery@imaginet.co.za

DATES:
Received: 15 Apr. 2020
Revised: 15 May 2020
Accepted: 25 May 2020
Published: 29 Sep. 2020

HOW TO CITE:
Overy N. 'Pollute the bay and poison the people': A short history of the Green Point marine sewage outfall, 1882–1992. S Afr J Sci. 2020;116(9/10), Art. #8173, 7 pages. https://doi.org/10.17159/sajs.2020/8173

ARTICLE INCLUDES:
☒ Peer review
☐ Supplementary material

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

EDITOR:
Jane Carruthers

KEYWORDS:
Cape Town, pollution, ocean, health, sewerage

FUNDING:
South African National Research Foundation (grant no. 118754)

'Pollute the bay and poison the people': A short history of the Green Point marine sewage outfall, 1882–1992

The City of Cape Town in South Africa pumps 40 million litres of untreated sewage into the Atlantic Ocean from the Green Point outfall pipeline every day. This results in microbial and chemical pollution of the sea (including persistent organic pollutants), marine organisms and recreational beaches, breaching the City's constitutional commitment to 'prevent pollution and ecological degradation' and, in doing so, it fails to uphold the constitutional right to an environment for citizens that is not harmful to 'health or well-being'. This article explores how the decision to build this marine outfall was reached in 1895. It illustrates how narrow economic interests from the 1880s until today have driven the City's commitment to the Green Point outfall despite a long history of opposition from citizens and scientists and repeated instances of pollution and ill-health. The findings reveal how, rather than being the cost-saving option that the City has always claimed it to be, its maintenance has cost enormous sums of money. The story of the Green Point outfall is one in which unimaginative, short-term monetary thinking has thwarted the search for an ecologically and hydrologically sustainable alternative means of sewage disposal – a legacy the City's residents and the oceans that surround it live with today.

Significance:

The findings show that the disposal of untreated sewage into the sea at Green Point has always been controversial (scientifically, politically and socially), and has resulted in episodes of illness among Cape Town citizens. It has never been the most economic option as repeatedly asserted by the City Council. The findings also illustrate that the City Council has consistently rejected ecologically and hydrologically sustainable alternative disposal options, and has denied problems associated with the outfall, despite overwhelming evidence to the contrary, a pattern which persists. These findings are important for civic governance, desalination, water quality, environmental humanities, health (human and marine), sanitation and urban planning.

The South African Constitution states that 'everyone has the right to an environment that is not harmful to their health or well-being' and legislative measures must be taken to 'prevent pollution and ecological degradation'.[1] The environmental legislative measures that govern South Africa's oceans are the *National Environmental Management Act (1998)* and the *National Environmental Management: Integrated Coastal Management Act (2004)*, read together with various supporting regulations, and coastal water quality guidelines issued by the South African Department of Water Affairs. This regulatory regime focuses

> *on maintaining or achieving water quality such that the water body remains or becomes fit for all designated uses … aquaculture, recreational use, industrial use, as well as the protection of biodiversity and ecosystem functioning.*[2]

In addition, South Africa is a signatory to various international agreements governing water quality, one of which is the Stockholm Convention on Persistent Organic Pollutants, that aims to protect human health and the environment from chemicals that remain intact in the environment for long periods.[3]

Despite the existence of this regulatory regime, it is clear that the discharge of untreated sewage from the Green Point sewage outfall negatively impacts water quality, damaging the marine environment and threatening human health. Research reveals high levels of microbial pollution, and the existence of persistent organic pollutants (both chemical and pharmacological) in seawater, and bioaccumulated in marine organisms.[4-6]

The Cape Town City Council has denied these findings, insisting that there is no evidence of pollution and thus no risk to either human or marine health.[7,8] This despite its own commissioned research which found evidence to the contrary.[9] What is revealing about this denial is how it replicates Council's response to similar complaints over almost a century concerning pollution and health risks caused by the Green Point outfall. As the following account shows, the Council has ignored or repudiated evidence of the dangers of the outfall from the 1880s, and yet has been forced to take some kind of remedial action time and again, when these dangers have proved genuine.

Giving evidence before a Parliamentary Select Committee on the sanitary state of Cape Town in 1888, H. Saunders, a City medical doctor, declared that sanitary conditions were 'as bad as any town in Christendom'.[10] Indeed, that the colonial government had required a Committee is evidence of its frustration with the City Council's apparent reluctance to do anything substantive about the state of sanitation in Cape Town.

The Council at the time was dominated by property owners and members of the so-called 'Dirty Party' who had resisted previous calls for sanitary reform fearing rates increases.[11] Pressure for reform came not only from the colonial government, however, but also from medical professionals concerned about high death rates (compared to English towns), and from merchants concerned that inadequate sanitation threatened the growth of the city by discouraging visitors and trade.[10,11(p.53-57),12]

However, in the late 1880s, the balance in the City Council fell to the so-called 'Clean Party', and thus ushered in the beginnings of meaningful sanitary reform. The chief objective of the new Council was to introduce a main drainage

system. In April 1888, the Council sought advice from E. Pritchard, a noted civil engineer from Birmingham, who drafted comprehensive main drainage plans. One Council member, T. Anderson, stated that all options should be considered and that sewage should not be simply 'thrown into our beautiful bay'.[13]

Pritchard published his report, 'Cape Town Sewerage' the following year. Describing the current system of sewage management 'as not only sickening in the extreme, but absolutely injurious to health', he proposed two drainage schemes.[14] The first, an outfall at the mouth of the Salt River had an estimated cost of £120 000, while the second, involving 'broad irrigation', would cost £200 000, with 'higher working expenses'. Due to its lower costs, he recommended the outfall.[14(p.8-10)]

The Council chose the outfall and allocated £85 000 to its construction. Because of the lower budget the outfall had to be located closer to the City near Fort Knokke, Woodstock. Numerous Councillors objected vociferously. Councillor J. Combrinck noting:

> *The health of the inhabitants of Cape Town and the purity of the Bay is the object I have in view; and I look confidently for the support and assistance of every man who has like interests at heart in my endeavour to warn the people of the danger of life which must inevitably follow the wholesale deposit of filth in Table Bay.*[15]

Councillor C. Goodspeed noted with regret that the Council had decided 'to carry out a portion of Mr Pritchard's scheme which has been, by amateurs, derailed'. He argued that the outfall should not be at Salt River, but 'in an entirely different direction, say Mouille Point'.[16]

The Council lost the vote even to raise £85 000, and by the end of 1890 the municipality was no closer to resolving its sanitation crisis.[17] In apparent exasperation, the colonial government intervened and contracted another drainage expert to consider the problem. In May 1891, C. Dunscombe, previously City Engineer of Liverpool, published his 'Report on the Sewerage of the City of Cape Town and the Disposal of its Sewage'.[18]

Dunscombe also proposed two schemes, largely similar to Pritchard's. Scheme One was an outfall at the Salt River Mouth, while Scheme Two was 'broad irrigation in combination … with a small amount of intermittent irrigation' on land outside Cape Town. The only difference between Dunscombe's and Pritchard's schemes was that Dunscombe recommended a second smaller outfall at the 'Green Point Lighthouse' to service Green Point and Sea Point.[19] He estimated that both sea outfalls would cost £119 000, while the broad irrigation scheme would cost £143 000.

Dunscombe concluded:

> *the sewage from the city can be safely and most economically disposed of by means of a free sea outfall as proposed in Scheme One … if purification of the sewage was the first consideration then scheme 2 could only be adopted, notwithstanding its extra cost for works and annual charges.*[19(p25)]

The *Cape Argus* campaigned strongly for Scheme 2, noting:

> *The sea outfall would be the cheaper scheme, but we say that even if we had the only assurance which any scientific man could give us, that in all human probability the Bay would never be polluted, we should prefer to feel that the sewage was conveyed to land … The cost will be but a few thousands more … the sentiment, if you will, against the further pollution of the Bay is so strong that we hope the land scheme will be decided upon.*[20]

Early in August 1891 the Council unanimously agreed to pursue Scheme 2 at a cost of £162 000. The mayor, De Villiers Graaff, said that the scheme would result in 'the preservation of the Bay, the purification of the city, and the lengthening of the lives of the citizens'. The *Cape Times* reported that the decision was met with a 'universal shower of congratulations which continued the whole day through', observing that the mayor believed that an outfall was 'a suicidal policy' which would 'pollute the Bay and poison the people'.[21]

However, 2 years after the Council had approved the scheme, it was discovered that land that had been allocated for it was unsuitable and the 'sewage farm' would have to be located further from the City, thus resulting in increased costs. By September, those costs had risen to £280 000 and Dunscombe informed the mayor that he could no longer recommend the 'sewage farm'.[22]

The City's Public Works and General Purposes Committee wanted Dunscombe to be fired because of inflated costs and implementation delays.[23] At the same time, this Committee ordered a survey of the ocean currents in Table Bay. Dunscombe was released from his contract and a newly appointed City Engineer, W. Olive, was 'as a matter of special urgency' given 'the very difficult problem of the disposal of the sewage of the City'.[24]

Olive produced a new drainage report in 1895, stating that 'the result of my deliberations are distinctly averse to any mode of dealing with the sewage other than by a sea outfall', dismissing any irrigation scheme as 'impracticable on account of great expense'.[25] After reviewing the ocean survey, he chose Green Point because 'the probabilities are highly favourable of the sewage being carried Northward' and because 'rough weather, and the churning action of the waves would aerate or oxygenate the sewage tending to render it innocuous'.[25(p.30-31)] Olive estimated the cost of a 182 m Green Point outfall at £235 000.[26] Funds were duly appropriated, and it was completed in 1905.[27]

There are a number of reasons why £235 000 was allocated in 1895, whereas only £85 000 had been rejected just 5 years earlier. Firstly, and perhaps most importantly, Cape Town experienced a severe economic recession in the late 1880s which dramatically reduced municipal revenue. This recession was, however, followed by an economic boom in the early 1890s which resulted in greatly increased municipal revenues. Municipal revenue in 1890 was £73 611, and in 1894 it totalled £498 963.[11(p.60),12(p.497)] Secondly, the 1890 plan was deeply unpopular because it proposed dumping sewage very close to the City centre at Woodstock, whereas the 1895 plan proposed Green Point which was, in 1895, far away enough to not even be considered part of the City of Cape Town. Lastly, the evidence indicates that by 1895 the point had finally been reached when something substantive had to be done about the state of sanitation in the City. It is clear that, by this time, the majority of City leaders, the media and the public had had enough of the delays and obfuscations of the past.

However, within just 6 years, there were numerous grievances from residents complaining that the outfall was a 'vile nuisance' because of a 'stench' that emanated from it so repugnant that they considered taking legal action.[28] On inspection, the City's Sanitary Superintendent agreed, finding that the 'atmosphere was very obnoxious' and that there was 'a quantity of sewage on the rocks and the lower part of the beach' near the outfall.[29]

The Council responded by spending a mere £150 on three 'ventilation shafts'.[30] However, the shafts made the smells even worse, and in late 1913 residents began subscribing to a legal fund.[31] After some procrastination, the shafts were demolished in 1916 as being useless.[32]There were more complaints 4 years later. These complaints followed a similar pattern, but now featured greater concern with sewage that was appearing on the beaches. So bad had the situation become by 1924 that the Cape Publicity Association contacted the City's Medical Officer of Health (MOH) and demanded that something be done about the 'the contamination of the sea with sewage' which is 'on many occasions visibly loaded with faecal matter'.[33]

In response, the MOH completed a report[34] which found evidence of 'drifting of sewage along the foreshore and the deposition of solid matters on to rocks' from both the Green Point and Sea Point outfalls. He argued that this pollution was 'dangerous' and drew attention to the

'exceptional amount' of enteric fever in Mouille Point and, to a lesser extent, Sea Point. The report included a table (Table 1) and observed:

> *… an examination of the foregoing table shows a striking excess of Enteric Fever in Mouille Point, an incidence rate for the five years amongst residents of the district being more than three times the rate for the whole Municipality and more than four times the rate in Cape Town.*[34]

Table 1: Incidence rate of enteric fever (number per 1000 per year) as per Higgins[34] report

Year	Europeans only		All races	
	East of Lighthouse	South West of Lighthouse	East of Lighthouse	South West of Lighthouse
1919–20	8.5	0.0	7.4	0.0
1920–21	21.4	0.0	18.4	0.0
1921–22	4.3	2.4	3.7	2.0
1922–23	14.9	7.1	14.8	6.1
1923–24	8.5	2.4	12.9	2.0
5-year average	11.5	2.4	11.4	2.0
Population	468	421	542	493

The MOH pointed out that the Green Point outfall received sewage from the Infectious Diseases Hospital located in that area. He believed that this probably accounted for the high incidence of enteric fever, especially to the east of the outfall because 'much more stercus is usually found to the east of the outfall, apparently owing to the circumstances that the surf tends to cast in on to that side.'

He concluded that the 'pollution of the foreshore cannot but be regarded as a possible cause of this prevalence of disease'. As a preliminary measure, he recommended treating sewage at the hospital before its release into the sewerage system, but left longer-term solutions to the City Engineer.[34]

The City Engineer, L. Davies, recommended that sewage from the hospital be chlorinated; by December 1925 it was, but complaints did not cease.[35] Two years later, a further MOH report noted that between March 1926 and March 1927 there had been five cases of enteric fever to the east of the outfall (equivalent to an incidence of 9.2 per 1000). It stated:

> *… the danger of enteric fever being caused by the sewage pollution of the foreshore and bathing places would exist even if a hospital for the treatment of typhoid patients did not drain into the sewer, and the chlorination of the hospital sewage will not remove that danger.*[36]

Upon inspecting the sea and foreshore, the MOH found a 'considerable quantity of solid excremental matter and other floating sewage matter' and appealed to the Council to act. On this occasion, the MOH felt compelled to make a forthright recommendation, arguing that the 'surest' way to solve the problem would be to cease discharging sewage into the ocean, or failing that, to treat the sewage before its release.[36]

Nearly 3 years after the MOH's first report, the City finally took action, asking Davies to recommend solutions. In June 1927, his report gave four options[35]:

1. To abandon the outfall, except for stormwater overflow purposes, and to pump the sewage to the Southern Suburbs Sewage Farm on the Cape Flats.
2. To purify the sewage by means of tanks and filters or an activated sludge plant before discharge into the sea through the existing outfall.
3. To sterilise the sewage by chlorination before discharge into the sea through the existing outfall.
4. To extend the outfall sewer and discharge the sewage deeper into the water.

The first option would cost too much: £500 000. The second was dismissed because 'capital costs and annual working expenses would be too high', and the Green Point area was 'not suitable for works of this kind'. Option three was rejected because chlorination had, in his opinion, yet to demonstrate that it eliminated disease-producing bacteria.

Davies's recommendation was to extend the outfall because it was already in a 'favourable position' and if 'extended into deeper water I am of the opinion that there will be very little risk of pollution of the Foreshore'. He recommended closing the Sea Point outfall and redirecting it to Green Point. He offered four different options for extending the pipeline. One option was an extension of either 152 m (£40 000) or 275 m (£53 000) of cast iron pipes on the sea bed, or extensions over the same distances but placed on top of a concrete mole for £71 811 or £117 453, respectively. He suggested the concrete mole because it 'will possess a comparatively long life'.[37]

In 1928, the Council opted for the 152 m concrete mole. However, it was then discovered that the breakwater at the harbour was to be extended by 460 m. This precipitated another report from Davies because 'the Breakwater Extension will have some effect upon the drifts around Green Point'. Davies then recommended an extension of an additional 550 m, giving a total length of 730 m. A mole would no longer be required, but 'non-corrosive nickel-chrome steel' pipes would be attached to the ocean floor and a pumping station would be installed.[38]

The Council reacted also to petitions from residents. One, in late 1927 and signed by 203 residents of Mouille Point and Sea Point, was concerned that since the outfall's construction in 1905 'numerous complaints have been made by the residents to the Council but no steps have been taken to abate the nuisance'. The signatories demanded 'immediate steps to abate and put an end to this nuisance'.[39] Legal action was threatened by J. Yolland, a Mouille Point resident.[40]

The following year the MOH informed the Council that enteric fever at Mouille Point continued to be 'considerably in excess of the rest of the Municipality'.[41] The threatened legal action then came about and the matter was heard by the Cape Supreme Court in March 1929 where Yolland sought interim relief via an interdict to immediately close the outfall. The judge refused to grant an interdict because there was no ready alternative for the disposal of sewage, but he stated that had there been an alternative, it would have been granted. The judge recommended that Yolland seek permanent relief in the High Court and he warned the Council:

> *It seems to me upon these affidavits the applicant has made a stronger case – I think a very much stronger case – than the respondent, and in going to action the respondent Council will very seriously have to consider what the result of that action may be, and they also have to consider their public responsibility in this matter.*[42]

Fearing further legal action (and expense) the mayor pushed for a private settlement with Yolland.[43] In October a secret agreement was signed with Yolland who committed to 'assist the Council in discouraging any other claims so founded in view of the scheme of extension so about to be undertaken' and, in exchange, the Council covered his legal fees and paid £800 in 'damages'.[44]

The city began to take action and there were tenders for the outfall extension to 640 m (not the planned 730 m), because of a 'boulder field'. Two 46-cm 'Gargantua Disintegrators' would be installed as an 'additional safeguard'.[45,46] By July 1931, the new pipes, pumps and disintegrators were in place.[47]

In August 1932, the MOH observed only one case of enteric fever in Mouille Point during the previous year and noted 'this is interesting in view of the fact that throughout the year the new outfall has taken the place of the old'.[48] Clearly the new system was abating the health hazards.

It was only some 25 years later that the Council began receiving complaints again from residents of Mouille Point about sewage on the beaches and in the sea. The City's Principal Chemist found that sewage was leaking 'at about 50 and 100 feet [15 m and 30 m] from the shore' causing 'considerable pollution of the beach'.[49] It is apparent from the archival records that the pipes installed in 1931 had, not surprisingly, reached the end of their useful lives by then. There are records of numerous reports of leaks from the pipeline and associated pollution, so much so that in 1962 the City's Engineer's Department recommended renewing and extending the outfall.[50] Four years later, when nothing had been done, the new City Engineer, S. Morris, reported that the corroded pipeline would need replacement within five years.[51]

Aerial photographs also revealed a 'most unsatisfactory state of affairs' with evidence of a 'sewage bloom' following the shoreline which 'completely envelops Granger Bay'.[52] He told the Council's Works and Planning Committee that:

> *... unless the effluent is discharged under such conditions as are acceptable under the Water Act the Committee may even be faced with an interdict to cease discharge into the sea; this, your Committee may recollect, was precisely what led to the construction of the present outfall.*[52]

He asked that a 'complete oceanographic investigation' be undertaken[52], and this survey, completed in 1970, recommended the outfall be tripled in length to 1800 m[53].

Despite this information, a decade passed before 'Finnish consultants' drew up a new design for the outfall.[54] The City may have been prompted into action because pressure came also from the national Department of Water Affairs which, in early 1981, warned the Council that

> *... levels of copper, iron and zinc emanating from the Green Point sewer discharge are well in excess of acceptable limits and consequently pose a serious pollution hazard ... the construction of a proper pipeline should proceed with the utmost urgency.*[55]

A few months later, the City Engineer said that a new pipe design had been approved, and would cost R6.1 million.[56] The City Council contracted the Council for Scientific and Industrial Research's (CSIR), National Institute for Water Research (NIWR) to assess two other options: the diverting of sewage to the treatment plant on the Cape Flats, or constructing a sewage purification plant in Green Point. The NIWR rejected both these options because pumping sewage to the Cape Flats was 'completely uneconomical' while treatment was rejected because it would 'absorb large areas of the sports fields at Green Point'. Instead, the NIWR favoured merely extending the outfall to 2700 m on the condition that 3-mm macerators were installed.[57]

Construction began in 1985 and was completed the following year. The new City Engineer, J. Brand, reported that R13.3 million had been spent on a high-density flexible polyethylene pipeline, predicting that this would 'render a nuisance-free utility for at least 50 years'.[58]

However, in August 1989, Brand informed the Council that that the pipeline had 'moved' after a storm.[59] In fact, the pipeline was so badly damaged that it had to be severed 280 m from the shoreline. This immediately resulted in extensive sewage pollution from Green Point to Granger Bay, forcing the shoreline to be closed to the public. Remarkably, however, Brand noted that 'surf faecal coliform counts' in the closed area were at 'levels roughly similar to those existing prior to the commissioning of the new outfall sewer at the beginning of 1986'.[60] Quite why closure was deemed necessary in 1989 but not in 1986 is unclear.

The public reacted angrily to the severing of the pipeline and resisted its repair or replacement, demanding that a solution be found that did not result in the dumping of untreated sewage into the sea.[61] Reflecting this public mood, the *Cape Times* reported:

> *The City Council's befouled chickens have come home to roost. In this day and age it should simply not be pumping raw sewage into the sea ... How much longer must Capetonians put up with the Council's casual attitude towards the pollution?*[62]

The Council thus commissioned new research on sewage disposal, budgeting R400 000 for the CSIR to investigate sea disposal, and R100 000 for the engineering consultants Ninham Shand to investigate land disposal.[63] The *Cape Times* questioned how seriously the Council was considering other options, observing that the chair of the Utilities and Works Committee had already publicly stated that the sea outfall was 'the most economically viable alternative'. The newspaper noted:

> *That's what they said before building the one that has just disintegrated. Will the Council ever learn? It has already wasted R13.5 million ... the estimated cost of the new pipeline is R20 million ... Persisting in the belief that the sea is the 'cheapest' alternative for the disposal of sewage is a short-sighted economy.*[64]

Councillors, when presented with research results from the CSIR and Ninham Shand in May 1990, expressed their concern that 'there was an element of bias, in favour of the marine option[65], while the City Engineer warned a public meeting that the outfall would mean a 2.2% increase in property rates, compared to 4.6% for land treatment[66]. It was clear that increasing rates was a concern.

To deflect accusations of bias, the Council contracted other engineering and environmental consultants, Kapp Prestedge Retief, to review the research findings and it also found in favour of the marine outfall on the basis of cost and because it believed that a fully functioning 1700-m outfall posed no danger to human or ocean health. It rejected Ninham Shand's option of a treatment plant at Green Point on the basis that the area was unsuitable for it. Kapp Prestedge Retief averred that the only alternative was land treatment at Paarden Island, at an estimated cost of R68 million. This cost was contrasted with reconstructing the outfall at R30 million:

> *Well-designed marine outfalls have long been considered as an efficient method of reducing sewage through dilution to levels of concentration equal to normal background levels found in the sea. Such systems have low capital and operating costs and involve virtually no terrestrial impacts such as odour or demand on space.*[67]

In August 1990, the Council budgeted the required R30 million for the new outfall, which was completed in 1993 and this is the outfall that currently pumps 40 million litres of untreated sewage into the ocean every day.

This account of the long history of the Green Point outfall reveals the Council's consistent approach to the disposal of sewage over many decades. At no point since its completion in 1905, has the Council seriously considered an alternative approach. The engagement with alternatives that has taken place, has been largely performative, more aimed at shaping public opinion in favour of the outfall, than in asking fundamental questions about how the City might be innovative in disposing of its sewage.

There is little doubt that technological lock-in has featured from the original decision to build the outfall in 1895. As sunk costs in the outfall, the pipe and the pumping station accumulated, it became increasingly difficult to abandon it in favour of alternatives. This situation has been exacerbated by the ever-increasing scarcity and high price of suitable land necessary for potentials like 'sewage farms'. And every time the Council took a decision to invest in renewing the outfall, it became less likely that an alternative scheme would ever be implemented.

This lock-in partly accounts for the Council's decisions always to choose the outfall as the 'cheapest' option. This bureaucratic decision-making is a by-product of the requirements of the political cycle which creates a perverse incentive to minimise costs in the shorter term, often at the expense of increased costs in the future.

The Council has also defended its unwillingness to spend more on alternatives by denying that the outfall is a problem, thus implying that alternatives – particularly on the grounds of public health – are unnecessary. This pattern of denial can be traced back to 1888 when the mayor told the Select Committee investigating the sanitary state of the City that 'matters are not as bad' as the medical evidence had suggested.[10]

The Council had blamed odours emanating from the outfall in the 1910s on putrefying seaweed, until multiple complaints came from lighthouse keepers at Mouille Point. Only then did the Council take action.[68,69] It is revealing that during the period this problem was being publicly denied, the City Engineer was privately writing to his fellow engineers in Durban and Port Elizabeth asking them how they abated their sewage smells.[70]

The situation was unchanged by the 1920s when the MOH reported on the high number of enteric fever cases and drew attention to the sewage regularly seen on the shore as a cause. City Engineer Davies stated then that accounts of visible sewage were 'exaggerated' and its link to enteric fever unproven. However, he also noted that teams of labourers were employed daily to collect stercus in buckets from the very same shore.[71]

This same narrative was to play out again in 1990 when the City Engineer stated, just before another R30 million was to be spent on the outfall:

> *Having taken all matters into account I conclude that a long marine outfall concept for the disposal of sewage has proved to be successful at Green Point for many years. There is no accumulation of pollutants and dispersal is totally effective with a consequent minimal health risk.*[72]

Another feature of this narrative has been ongoing opposition to the outfall not only from residents, but also from sanitation, engineering, and health professionals. Some of these include condemnation from Cape Town's Water Engineer, who argued even before the outfall was constructed that disposing sewage into the sea, while cheap, led to the contamination of 'sea life' and threatened 'public health'.[73]

But almost at the start of this saga, in 1911, A. Snape, then Professor of Civil Engineering at the South African College (later to become the University of Cape Town), believed that it was a 'pity that the whole scheme was not designed for the sewage to go to the Flats'. He thought that if there really was no alternative to sea disposal, there should be 'some form of tank operation for dealing with the raw sewage before it was discharged into the sea'.[27(p.76-77)]

In 1933, writing in *Minutes of the Proceedings of the South African Society of Civil Engineers*, E. Croghan, Cape Town resident and noted analytical chemist and expert on sanitation and sewage disposal, observed that the time had come to chlorinate all sewage entering the sea to protect bathers and those who collected shellfish. He specifically drew attention to Melbourne in Australia, which had a population four times the size of Cape Town but did not drain its sewage into the sea.[74]

A few years later, H. Wilson, the biochemist for Johannesburg, lamented the disposal of raw sewage into the sea, suggesting that towns go through a number of phases before they realise that it is not viable. Wilson contended that it was a shame that 'the same mistakes were going on in South Africa as have already been made in other parts of the world'.[75] In 1961, O. Coetzee, senior bacteriologist of the CSIR's NWIR, writing in the *South African Medical Journal*, observed the 'almost irresistible temptation of seaside municipalities' to dispose of sewage into the sea, despite it being 'bacteriologically objectionable'.[76]

A further feature of the Council's approach to sewage disposal has been that no matter how 'objectionable' sewage might be as a characteristic of urbanisation, disposing of it in the sea appeared to solve the problem by making sewage disappear. Throughout the history of the Green Point outfall, the Council has trusted in the ability of the sea to sanitise sewage. But there was a dilemma, as expressed by Olive in 1895, that while the outfall 'would cause no nuisance', at the same time, 'it is pretty generally accepted that raw sewage cannot be disposed of anywhere without risk of nuisance … it is certain that at some future time it will make its presence felt'.[77]

The very act of always extending the outfall reveals the same dissonance. It is reassuring because it distances citizens yet further from the sewage by disposing of it further out to sea. However, the constant need to extend the outfall because sewage still finds its way to the shore, reveals how futile is the very act of extension. The time has come for the Council to think differently about sewage and the sea. The multiple entanglements of the human, the natural, and the technical need to be acknowledged and faced beyond the binaries of technologists, or the parsimony of the City's financial managers and ratepayers. There is also the recognition today that untreated chemical and pharmacological pollutants produce environmental 'slow violence' as they persist in the ocean and make their way into marine life, and eventually back into humans.[78] Only by facing these manifestations of the Anthropocene with openness and a receptivity to new challenges and ways of overcoming them, will the Council meet its Constitutional commitments to maintain a healthy environment for all living species.

Acknowledgements

I would like to acknowledge the assistance of Prof. L. Petrik and Prof. L. Green in commenting on earlier drafts of this article. I also gratefully acknowledge the support of the National Research Foundation (South Africa). This work is based on research supported wholly by the National Research Foundation of South Africa (grant number 118754) – South Africa–Norway cooperation on ocean research including blue economy, climate change, the environment and sustainable energy (SANOCEAN).

Competing interests

I declare that there are no competing interests.

References

1. South African Constitution. Section 24a & b(i); 1996.
2. South African water quality guidelines for coastal marine waters – Volume 1: Natural environment and mariculture use. Pretoria: Department of Environmental Affairs; 2018. p. xxiii.
3. Stockholm Convention on Persistent Organic Pollutants [homepage on the Internet]. No date [cited 2020 May 14]. Available from: http://www.pops.int/
4. Petrik L, Green L, Abegunde A, Zackon M, Sanusi Y, Barnes J. Desalination and seawater quality at Green Point, Cape Town: A study on the effects of marine sewage outfalls. S Afr J Sci. 2017;113(11/12), Art. #a0244, 10 pages. https://doi.org/10.17159/sajs.2017/a0244
5. Zackon M. Public protest and the production of evidence regarding Cape Town's marine sewer outfalls [honour's dissertation]. Cape Town: University of Cape Town; 2017.
6. Cape Town Outfalls Monitoring Program: Surveys made in 2015/2016. CSIR Report; 2017 June.
7. Harvey J. 'No human health risks' from sea marine outfalls. Sentinel News. 2017 November 24 [cited 2020 Apr 06]. Available from: www.sentinelnews.co.za/news/no-human-health-risks-from-sea-marine-outfalls-12130629
8. City of Cape Town. Academics' unsubstantiated bombshells trigger panic free of accountability'. Daily Maverick. 2019 March 22 [cited 2020 Apr 06]. Available from: www.dailymaverick.co.za/article/2019-03-22-academics-unsubstantiated-bombshells-trigger-panic-free-of-accountability
9. Cape Town Outfalls Monitoring program: Surveys made in 2015/2016. CSIR Report; 2017 June.
10. KAB (Cape Town Archives Repository) HA (House of Assembly) 746(a) 112 Report of the Select Committee on Sanitation in Cape Town; 1888.
11. Bickford-Smith V. Ethnic pride and racial prejudice in Victorian Cape Town: Group identity and social practice 1875 – 1902. Cambridge: Cambridge University Press; 1995. p. 48.

12. Van Heyningen E. Public health and society in Cape Town 1880-1910 [unpublished PhD]. Cape Town: University of Cape Town; 1989. p. 64, 257.
13. The Drainage of Cape Town. Cape Times. 1888 April 17.
14. Pritchard E. Cape Town sewerage. Cape Town: Murry and St. Ledger; 1889. p. 8.
15. Letter Combrinck J. to the Editor. Cape Times. 1890 June 14.
16. Letter Goodspeed C. to the Editor. Cape Times. 1890 June 14.
17. Drainage Scheme. Cape Times. 1890 June 17.
18. Drainage in Cape Town. Cape Times. 1891 August 04.
19. Dunscombe C. Report on the sewerage of the City of Cape Town and the disposal of its sewerage. Cape Town: Richards and Sons; 1891. p. 6.
20. The Two Schemes. Cape Argus. 1891 June 22.
21. Drainage in Cape Town. Cape Times. 1891 August 04.
22. KAB 3/CT (Cape Town) 1/1/1/42 Report of the Public Works Committee on the Disposal of the Drainage of the City; 1893 Sept. 28.
23. KAB 3/CT 1/1/1/42 Special Meeting of the Public Works Committee; 1893 Dec. 12 and Meeting of the Public Works Committee, 1893 Dec 21.
24. KAB 3/CT 4/1/5/304 C1424/5 Green Point Outfall: Alleged Sewage Pollution of Beach: Three Anchor Bay. Sea Point: Extension of Outfall at Green Point. Minutes of a Special Meeting of the Town Council of the City of Cape Town', 1895 Dec. 16.
25. KAB 3/CT 4/1/5/304 C1424/5 The Corporation of the City of Cape Town Report of City Engineer and Surveyor Upon the Question of Sewage Disposal. 1895 July 26; p. 1, 3–8.
26. KAB 3/CT 4/1/5/304 C1424/5 Minutes of a Special Meeting of the Town Council of the City of Cape Town. 1895 Dec. 16.
27. Perkins H. The Cape Town Main Drainage Scheme. Minutes of the Proceedings of the South African Society for Civil Engineers. Vol. IX. 1911; p. 69. The much smaller Sea Point outfall was completed in 1901.
28. KAB 3/CT 4/4/1/41 A398 Ventilating Shaft, Green Point Common: Sewerage Nuisance, Green Point Common 1911-16. Letter Phillips to Verster J.; 1911 Nov. 11.
29. KAB 3/CT 4/1/2/19 A298/2 Sewer Outfall, Green Point, 1913. Letter Simpson Wells A. to Jeffries W. 1912 June 13.
30. KAB 3/CT 4/4/1/41 A398 Letter Jasper Anderson A. to Chairman and Members of the Public Health and Building Regulations Committee; 1911 Nov 01.
31. Mouille Point Drainage. Cape Times. 1913 December 31.
32. KAB 3/CT 4/1/2/19 A298/2 Sewer Outfall, Green Point, 1913. Letter Jeffries W. to Public Works and General Purposes Committee 1913 Jan. 8 and Letter Town Clerk to Jeffries W. 1916 Jan. 27.
33. KAB 3/CT 4/1/5/304 C1424/5 Letter Acting Secretary of the Cape Publicity Association to Dr Higgins; 1924 May 13.
34. KAB 3/CT 4/1/5/304 C1424/5 Higgins S. Report to the Health and Building Regulations Committee; 1924 August 15.
35. KAB 3/CT 4/1/5/304 C1424/5 Davies L. The Corporation of the City of Cape Town – Report on Disposal of Sewage form the Cape Town and Green and Sea Point Areas; 1927 June 15.
36. KAB 3/CT 4/1/5/304 C1424/5 Higgins S. Report to the Health and Building Regulations Committee; 1927 March 7.
37. KAB 3/CT 4/1/5/304 C1424/5 The Corporation of the City of Cape Town – Report on Disposal of Sewage form the Cape Town and Green and Sea Point Areas; 1927 June 15.
38. KAB 3/CT 4/1/5/304 C1424/5 Davies L. City of Capetown – Final Report on Disposal of Sewage from the Capetown and Green and Sea Point Area; 1929 March 12.
39. KAB 3/CT 4/1/5/304 C1424/5 Petition to the Mayor and Councillors of the City of Cape Town; 1927 December.
40. KAB 3/CT 4/1/5/304 C1424/5 Letter Stent and Simpson to Town Clerk; 1929 Dec. 12.
41. KAB 3/CT 4/1/5/304 C1424/5 Letter Higgins to Town Clerk; 1929 Dec. 28.
42. KAB 3/CT 4/1/5/306 C1425/5 Green Point Outfall: Alleged Sewage Pollution of Beach: JW Yolland versus City Council, 1929. Judgment in the Case in the Supreme Court of SA (Cape of Good Hope Division); 1929 March 29.
43. KAB 3/CT 4/1/5/304 C1424/5 Green Point Outfall: Alleged Sewage Pollution of Beach: Three Anchor Bay. Sea Point: Extension of Outfall at Green Point. Statement made by the Mayor to Council; 1929 Sept. 26.
44. KAB 3/CT 4/1/5/305 C124/5 Green Point Outfall: Alleged Sewage Pollution of Beach: Three Anchor Bay, Sea Point: Extension of Outfall at Green Point: Installation of Disintegrators 1930. Agreement Reached between Yolland and City of Cape Town; 1929 Oct. 4.
45. KAB 3/CT 4/1/5/304 C1424/5 Report Davies L. to Sanitation and Streets and Drainage Committees; 1929 Dec. 18
46. KAB 3/CT 4/1/5/304 C1424/5 Letter Town Clerk to Davies L.; 1930 Dec. 1.
47. KAB 3/CT 4/2/1/1/373 62/31 – Improvement of Green Point Sewage Outfall, 1931. Report Davies. L to Sanitation and Streets and Drainage Committees; 1931 Oct. 31.
48. KAB 3/CT 4/2/1/1/353 Erection of Extension of Sewage Equipment and Pumping Stations, Muizenberg and Green Point, 1932. Letter Higgins to Davies; 1932 August 29.
49. KAB 3/CT 4/2/1/5/49 RG59 – Green Point Outfall Divers Reports, Repairs 1946-67. Letter Abbot. R (City of Cape Town Principal Chemist) to Roads and Drainage Engineer; 1956 July 27.
50. KAB 3/CT 4/2/1/5/49 RG59 – Letter Nyman. C (City Engineers Department) to Harris T. (CSIR) 1962 Dec. 30.
51. KAB 3/CT 4/2/1/5/49 RG59 - Report Morris S. to Works and Planning Committee; 1966 Nov. 30.
52. KAB 3/CT 4/2/1/5/49 RG59 - Report Morris S. to Works and Planning Committee; 1967 Feb. 7.
53. KAB 3/CT 4/1/11/773 G65/6/275/48 Executive Committee – Cape Town City Council. Matters referred to Council: Green Point Outfall Sewer Report from Institute of Oceanography.
54. KAB 3/CT 1/4/2/9/1/29 City of Cape Town Utilities and Works Committee Sept 1980 – Aug. 1981 (Vol 1). Letter City Engineer to Utilities and Works Committee; 1980 Sept. 15.
55. KAB 3/CT 1/4/2/9/1/29 Letter Lusher J. (Department of Water Affairs) to Town Clerk; 1981 March 19.
56. KAB 3/CT 1/4/2/9/1/29 City of Cape Town Utilities and Works Committee Sept 1980 – Aug. 1981 (Vol 2). Letter City Engineer to Utilities and Works Committee; 1981 June 8.
57. KAB 3/CT 1/4/2/9/1/33 City of Cape Town Utilities and Works Committee Sept 1982 – Aug. 1983 (Vol 1). Cillie G. Review of Proposals for construction of a marine outfall sewerage at Green Point – Cape Town. National Institute for Water Research; no date.
58. Brand J. Green Point Marine Outfall Project. Civil Engineer; 1983 Sept. p. 516.
59. KAB 3/CT 1/4/2/9/1/46 City of Cape Town Utilities and Works Committee Apr. 1989 – Aug 1989. Minutes; 1989 August 7.
60. KAB 3/CT 1/4/2/9/1/47 City of Cape Town Utilities and Works Committee Sept. 1989 – March 1990. Letter City Engineer to Utilities and Works Committee' 1989 Oct. 26.
61. Fijen A, Marsden M, Green Point Sewage Disposal Options. Water, Science and Technology. Vol. 25, No. 9, 1992, p. 135 and KAB 3/CT 1/4/2/9/1/48 City of Cape Town Utilities and Works Committee March 1990 – Aug 1990. Final Report – Summary of Facts; 1990 August 13. p. 2.
62. Using the sea as a sewer. Cape Times. 1989 August 17.
63. KAB 3/CT 1/4/2/9/1/47 - City of Cape Town Utilities and Works Committee Sept. 1989 – March 1990. Minutes 1989 Oct. 27.
64. A chance to save the sea. Cape Times. 1989 September 02.
65. KAB 3/CT 1/4/2/9/1/48 Special Meeting of the Utilities and Work Committee and the Amenities and Health Committee. 1990 May 22.
66. Undersea sewage may mean 'cheaper' rates. Cape Times. 1990 May 24.

67. KAB 3/CT 1/4/2/9/1/48 Green Point Sewage Disposal: Review and Comparative Appraisal of Land and Marine Options, Kapp Prestedge Retief; 1990.

68. KAB 3/CT 4/1/2/19 A298/2 Letter City Engineer to Works Committee; 1913 Jan. 8

69. KAB 3/CT 4/1/2/19 A298/2 Letter Assistant General Manager South African Railways and Harbours to Town Clerk; 1913 Jan. 22.

70. KAB 3/CT 4/1/2/19 A298/2 Letter City Engineer to City Engineers in Durban and Port Elizabeth; 1912 Dec. 12.

71. KAB 3/CT 4/1/5/304 – C1424/5 Notes by City Engineer for Coulter C. 1929 Feb. 15.

72. KAB 3/CT 1/4/2/9/1/48 City of Cape Town Utilities and Works Committee March 1990 – Aug 1990. Final Report – Summary of Facts; 1990 August 13. p. 9.

73. Wynne-Roberts R. Sewage Disposal. Minutes of the Proceedings of the Cape Society of Civil Engineers; 1904. p. 10–12.

74. Croghan E. The Evolution of Modern Sewage Disposal. Minutes of the Proceedings of the South African Society of Civil Engineers; 1933. Vol. 31. p. 160–161.

75. Wilson H. Sewage disposal in South Africa. S Afr J Sci. 1937;XXXIII:225.

76. Coetzee O. Comments on sewage contamination of coastal bathing waters. S Afr Med J. 1961;35:261.

77. KAB 3/CT 4/1/5/304 C1424/5 The Corporation of the City of Cape Town Report of City Engineer and Surveyor Upon the Question of Sewage Disposal. 1895 July 26; p. 2, 13.

78. Nixon R. Slow violence and the environmentalism of the poor. Cambridge, MA: Harvard University Press; 2011.

Statistical classification of South African seasonal divisions on the basis of daily temperature data

AUTHORS:
Adriaan J. van der Walt[1,2]
Jennifer M. Fitchett[1]

AFFILIATIONS:
[1]School of Geography, Archaeology and Environmental Studies, University of the Witwatersrand, Johannesburg, South Africa
[2]Department of Geography, University of the Free State, Bloemfontein, South Africa

CORRESPONDENCE TO:
Jennifer Fitchett

EMAIL:
Jennifer.fitchett@wits.ac.za

DATES:
Received: 13 Nov. 2019
Revised: 10 June 2020
Accepted: 11 June 2020
Published: 29 Sep. 2020

HOW TO CITE:
Van der Walt AJ, Fitchett JM. Statistical classification of South African seasonal divisions on the basis of daily temperature data. S Afr J Sci. 2020;116(9/10), Art. #7614, 15 pages. https://doi.org/10.17159/sajs.2020/7614

ARTICLE INCLUDES:
☒ Peer review
☐ Supplementary material

DATA AVAILABILITY:
☐ Open data set
☐ All data included
☒ On request*
☐ Not available
☐ Not applicable
**from South African Weather Service*

EDITOR:
Yali Woyessa

KEYWORDS:
seasonality, classification, climatology, biometeorology, Euclidean cluster analysis

FUNDING:
University of the Free State; DSI-NRF Centre of Excellence for Palaeoscience

Across South Africa, a wide range of activities is influenced by differences in seasonality. In a South African context, there is little consensus on the timing of seasonal boundaries. Inconsistency exists through the use of ad-hoc approaches to define seasonal boundaries across South Africa. In this paper, we present one of the very first uniform statistical classifications of South African seasonal divisions on the basis of daily temperature data. Daily maximum and minimum temperature data were obtained from 35 selected South African Weather Service meteorological stations that had sufficiently complete data sets and homogeneous time series, spanning the period 1980–2015. An Euclidean cluster analysis was performed using Ward's D method. We found that the majority of the stations can be classified into four distinct seasons, with the remaining 12 stations' data best classified into three seasons, using T_{avg} as the classifier. The statistically classified seasonal brackets include summer (October/November/December/January/February/March), early autumn (April) and late autumn (May), winter (June/July/August), and spring (September). Exploring the boundaries of seasons, the start of summer and end of winter months follow a southwest to northeastwards spatial pattern across the country. Summers start later and winters end later in the southwestern parts of the country, whereas in the northeast, summers start earlier and winters end earlier.

Significance:

- The findings contribute to the common knowledge of seasonality in South Africa.
- New seasonal divisions in South Africa are proposed.

Introduction

Seasonal differences in climate, day length, and plant activity form the primary environmental control on a wide range of activities. These activities include economic and agricultural practices[1], which are affected by the length and timing of growing seasons[2,3], the related timing of sowing and harvest, and the necessity for irrigation and fertilisation. Climatic seasons also influence resource management and energy demand[4], tourism[5], social and economic planning[6], hydrology[1] and health[7]. The capacity to accurately determine the start and end of seasons is thus of critical importance. However, in the South African context, there is little consensus as to distinct seasonal boundaries. This lack of consensus is unusual. For most parts of the northern hemisphere[8,9] – including Italy[10], USA[11], China[12], Poland[13] and Finland[14] – seasonal boundaries are well established and clearly communicated.

South Africa is classified as a semi-arid country[15] which is situated in the mid-latitudes and subtropics[16,17]. The South African climate is influenced by major synoptic systems, including the semi-permanent subtropical high-pressure systems[16] and the variability of the Inter-Tropical Convergence Zone (ITCZ)[18-20]. The resultant continental anticyclones, ridging anticyclones, westerly waves, tropical easterly waves, and cut-off lows[16,21,22], produce a pronounced climatic seasonality across the country[20]. However, there is little consensus on the timing of seasons in South Africa, and approaches in defining seasonal boundaries often vary on the basis of the application (Table 1).[18] The South African Weather Service (SAWS) even highlights that there is no official designation and definitions of seasons.[23]

The most basic classifications commonly used are astronomical, meteorological and phenological divisions.[23] Astronomical summer is defined as the period from the summer solstice to the vernal equinox. By this classification, autumn is defined to conclude at the winter solstice, and winter spans the winter solstice to the spring equinox.[23,24] It is well known that the earth–sun geometry affects the seasons; however, there is no direct link between astronomical seasons and mean weather variations.[24,25] Meteorological classification refers to the subdivision of four equal-length periods of 3 months each, mostly and commonly used in the temperate latitudes.[23] In South Africa, the meteorological classification of temperature seasonality is the most widely used (Table 1).[2,26-29] Most agroclimatological studies use the conventional break of 3 months but may extend it to the farming season of the specific regions that are under investigation and use two distinct seasons, summer and winter (Table 1).[30] Climate modelling projections and analyses of influences of climatic factors studies often use six run-on seasons that coincide with synoptic circulations, with the latter coinciding with epidemiological seasons of heightened disease-risk.[31] Some climatologists classify seasons on an ad-hoc basis that is appropriate to specific regions, with classifications such as hot season, cold season, post-rainy season and growing season, with no direct relationship with calendar months. These seasons are defined for convenience and to suit the data output rather than to drive sensible analytical processes.[25-27] Phenological studies can also be used to define annual seasonality, and to reveal shifts in the timing of these events[32], as phenological shifts are often directly related to changes in local air temperatures[33].

Classification of South African seasons based on rainfall patterns is similarly complicated due to the variety of rainfall regimes, and thus likewise no standard definition has been adopted[18], and discrepancies exist between the seasonal brackets of rainfall and temperature-related classifications[34]. The differences are further complicated by the influence of distance from large water bodies, and the variation in heat from the Indian and Atlantic Oceans.[24]

Here we present one of the first statistical classifications of South African seasonal divisions on the basis of daily temperature data for 35 weather stations spanning the country. We argue that this method represents a more standardised and objective approach to the classification of seasons, particularly in a region that spans the subtropics and mid-latitudes.

Study region

South Africa is located within the latitudes 22–35°S and longitudes 17–33°E, and is bordered by the Atlantic Ocean in the west and southwest and the Indian Ocean to the south and southeast. It shares political boundaries with Mozambique, Zimbabwe, Botswana, Namibia and Eswatini (Swaziland), and encloses Lesotho (Figure 1).[35] The climate of South Africa, in particular temperature, is governed by the complex interaction between the subtropical location, the altitude of the interior plateau, the position of the subcontinent with respect to the major atmospheric circulation features, and the oceans on all sides except the north.[19,20,36] The subcontinent lies within the subtropics, with rainfall dominated by convective storms in the north and mid-latitude cyclones to the south.[16,37] The influence of the tropical and temperate pressure regimes, and the intra-annual migration of the inter-tropical convergence zone (ITCZ) results in pronounced seasonal differences in rainfall and temperature patterns over South Africa.[19] The ITCZ shifts with the monthly and seasonal changes of the sun's maximum insolation and the location of dominant atmospheric high- and low-pressure systems.[19,37] The high-pressure systems sit over the southern tip of the subcontinent in summer, and over the interior during winter. These high-pressure systems are interrupted by mid-latitude cyclones.[38] The influences of the subtropical high-pressure belt, and the mid-latitude westerlies with associated fronts vary significantly inter- and intra-annually over the subcontinent.[38] These interactions between tropical and temperate disturbance have significant consequences for the weather of the subcontinent.[16] The orography of South Africa influences the temperature distribution over the country such that the escarpment forms a climatic division between the high plateau and the low-lying coastal regions in the east and southeast (Figure 1).[19] The southern and eastern escarpments are the regions with the lowest temperatures, due to the decrease in temperature with altitude.[39,40] The oceans surrounding South Africa influence the temperatures experienced along the coastal areas.[39,40] The Indian Ocean, on the east, is warmed by the western boundary Agulhas Current, while the Atlantic, on the west coast, is cooled by the eastern boundary Benguela Current (Figure 1).[19,39] All these factors result in a broad east–west temperature gradient, with the Northern Cape experienceeing the lowest rainfall and highest temperatures in the country.[39,40]

Table 1: Published temperature seasonality classifications for South Africa

Publication	Research topic	Seasonal temperature brackets	Seasonal temperature brackets approach
Karl et al., 1993[26]	Daily maximum and minimum temperatures	Dec/Jan/Feb (summer) Mar/Apr/May (autumn) Jun/Jul/Aug (winter) Sep/Oct/Nov (spring)	Three-month mean
Klopper et al., 1998[27]	Seasonal maximum temperature predictability in South Africa	Dec/Jan/Feb (summer) Mar/Apr/May (autumn) Jun/Jul/Aug (winter) Sep/Oct/Nov (spring)	Climatological austral seasons
Craig et al., 2004[31]	Malaria case data in KwaZulu-Natal	Nov/Dec/Jan/Feb/Mar (summer) Mar/Apr/May/Jun (autumn) Jun/Jul/Aug (winter) Aug/Sep/Oct/Nov (spring)	Defined approach
Kruger and Shongwe, 2004[28]	Temperature trend in South Africa	Dec/Jan/Feb (summer) Mar/Apr/May (autumn) Jun/Jul/Aug (winter) Sep/Oct/Nov (spring)	Subdivided into four seasons
Benhin, 2006[30]	Climate change and South African agriculture	Dec/Jan/Feb/Mar/Apr/May (summer) Jun/Jul/Aug/Sep/Oct/Nov (winter)	Mean monthly temperatures; farming seasons
Tshiala et al., 2011[2]	Analysis of temperature trends in Limpopo	Dec/Jan/Feb (summer) Mar/Apr/May (autumn) Jun/Jul/Aug (winter) Sep/Oct/Nov (spring)	Defined following the usual conventions
Lazenby et al., 2014[7]	Seasonal temperature prediction over southern Africa and human health	Sep/Oct/Nov; Oct/Nov/Dec; Nov/Dec/Jan Dec/Jan/Feb; Feb/Mar/Apr (summer to late summer)	Southern Africa is mainly controlled by influences from the tropics
Kruger and Nxumalo, 2017[29]	Temperature trends in South Africa	Dec/Jan/Feb (summer) Mar/Apr/May (autumn) Jun/Jul/Aug (winter) Sep/Oct/Nov (spring)	Mean monthly temperatures
Tshiala and Olwoch, 2010[56]	Tomato production in Limpopo	Dec/Jan/Feb (summer) Mar/Apr/May (autumn) Jun/Jul/Aug (winter) Sep/Oct/Nov (spring)	Mean monthly temperatures

Data and methodology

For this study, daily maximum and minimum temperature data were obtained from 35 selected SAWS meteorological stations (Figure 1; Table 2) that had a minimum of 30 years of data, sufficiently complete data sets and homogeneous time series, spanning the period 1980–2015. These stations were selected as they span the country, ranging from 22°S to 35°S and 15°E to 33°E with an intended 1° interval (Figure 1). Before performing any statistical techniques, exploratory data analysis was applied to investigate the data homogeneity due to inevitable changes in aspects including observation sites, station relocation, observation practices/procedures and time.[29] However, in the context of the study, sudden increases or decreases in values over a prolonged period would not have significantly influenced the results. Visible outliers in the data series were checked by comparison with data from surrounding stations spanning the period of interest as well as reports of anomalous weather in the media.

Table 2: Information on selected meteorological stations

Province	Station name	Climate number	Latitude	Longitude	Altitude (m.asl)
Eastern Cape	East London	00595729	-33.03	27.83	124
	Dohne	00798116	-32.52	27.47	899
	Mthatha	0127272A4	-31.55	28.67	745
	Port Elizabeth	00351795	-33.98	25.60	60
Free State	Bloemfontein	02615161	-29.10	26.30	1354
	Bethlehem	03315859	-28.24	28.33	1689
	Welkom	03643001	-27.99	26.66	1344
Gauteng	Johannesburg Int	*04763990*	-26.14	28.23	1706
	Zuurbekom	04755288	-26.30	27.81	1580
KwaZulu-Natal	Cedara	02394820	-29.54	30.27	1071
	Durban	0240808A2	-29.97	30.95	96
	Escourt	03006901	-29.01	29.86	1144
	Shaleburn	02376184	-29.80	29.35	1614
Limpopo	Mara	072200991	-23.14	29.56	894
	Warmbad Towoomba	05895941	-24.90	28.32	1143
	Polokwane	0677802BX	-23.84	29.45	1226
Mpumalanga	Skukuza	05961793	-24.99	31.59	276
	Carolina	04801547	-26.07	30.11	1667
Northern Cape	Calvinia	0134479A3	-31.48	19.76	975
	De Aar	01698801	-30.66	23.99	1286
	Fraserburg	01130258	-31.91	21.51	1264
	Kimberley	02904684	-28.80	24.77	1198
	Pofadder	0247668A4	-29.12	19.38	984
	Springbok	0214700B2	-29.66	17.87	1006
	Sutherland	00882931	-32.38	20.67	1459
	Upington	03174765	-28.43	21.27	836
	Vanwyksvlei	01935613	-30.35	21.82	965
North-West	Vryburg	04322383	-26.95	24.65	1245
	Mahikeng	05080470	-25.80	25.54	1280
Western Cape	Cape Agulhas	00030204	-34.83	20.01	11
	Cape Columbine	00606209	-32.83	17.86	62
	Cape St Blaze/Mossel bay	00122517/0012215	-34.18/-34.19	22.15/22.13	135
	Cape Town	0021178A3	-33.96	18.60	42
	Vredendal	0106880A2	-31.67	18.50	42
	George	00126617	-34.00	22.38	197

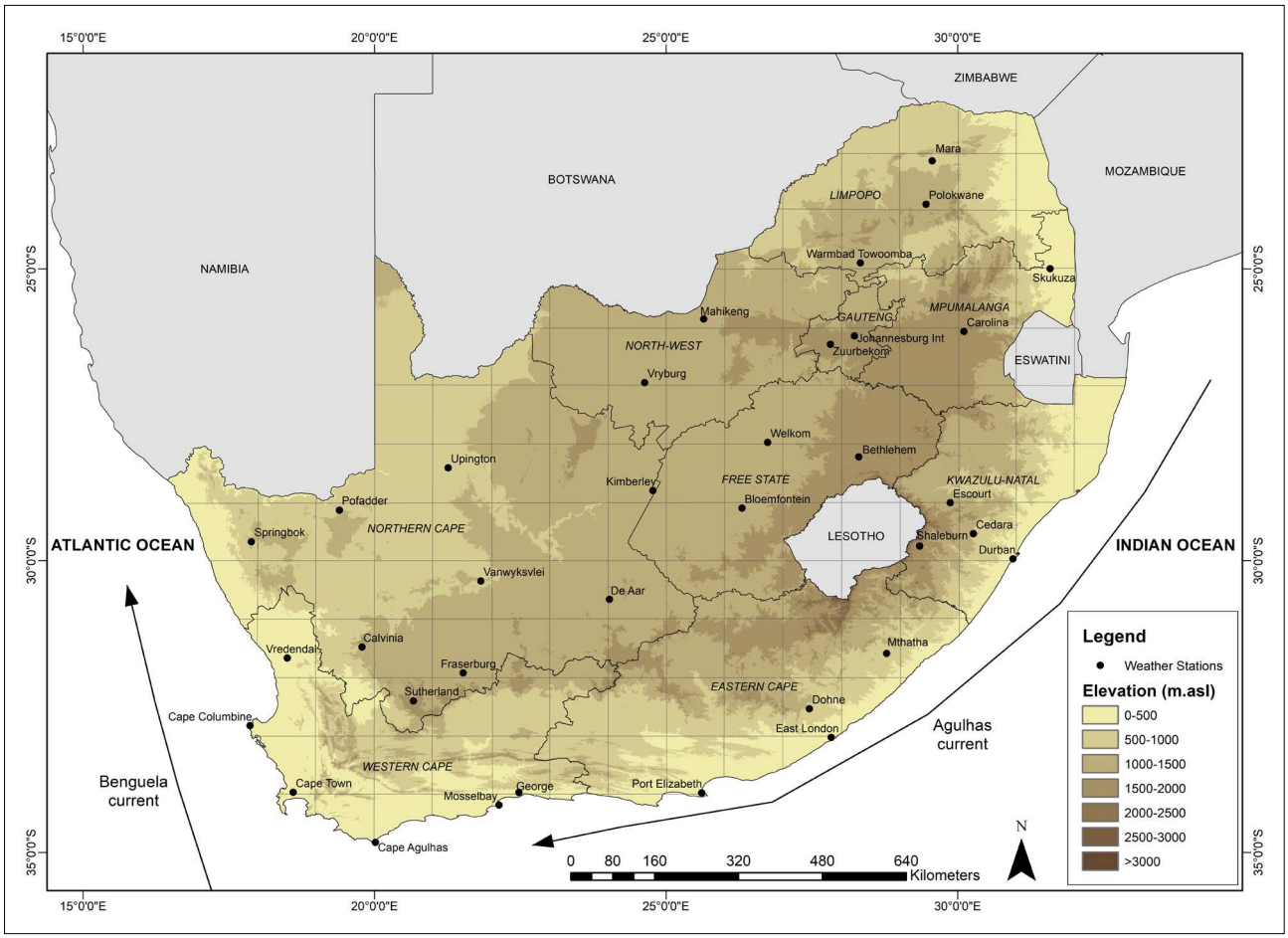


Figure 1: Map of the study region of South Africa, indicating 1° intervals used to select the meteorological stations across South Africa, the topography, the surrounding oceans and their respective ocean currents.

The data sets of the selected stations were subjected to quality control. As a first step, all dates and times were checked, and two decimal point rounding was used to maintain consistency throughout. Missing weather station data were replaced with data from a station adjacent to the site within a 10-km radius, or, if not possible, replaced with the 5-day running average. If data were not available for more than five consecutive days, that period was excluded from the analysis.

Cluster analysis

Cluster analysis was performed using Ward's D method, defined by the Euclidean distance between variables, utilising the cluster, vegan and rioja packages in R.[41-43] Euclidean cluster analysis was initially supervised at four seasonal divides and validated by using the dendogram package average silhouette width (ASW) calculation.[44] The ASW value measures the degree of confidence in between-group distances and strength of within-group homogeneity.[45] If not significant, two, three, five and six seasonal divides were used serially until the cluster was significant, using orders of magnitude put forward by Kaufman and Rousseeuw[46] as reference for measures. The ASW was calculated, together with the cophenetic correlation coefficient (CPCC) for interpretation, evaluation and validation of consistencies within the cluster and groupings.[42,47] The CPCC measures the correlation between the original pairwise distance matrix and the cophenetic distance matric of the dendrogram. This allows for the verification of the quality of the grouping.[47-49] The closer the cophenetic correlation coefficient is to a value of one, the better the grouping quality.[49] The cluster analysis results for maximum (T_{max}), minimum (T_{min}) and average (T_{avg}) temperatures are given in Table 3. To investigate the spatial patterns, the cluster analysis outputs, and start and end dates of summer and winter, were spatially interpolated using the Inverse Distance Weighted (IDW) method using ArcGIS software.[50] It has been found that IDW interpolates station data accurately.[51] Additionally, annual mean graphs were produced for each of the temperature metrics.

Results

Cluster analysis

Results will mainly focus on T_{avg}, with reference to T_{max} and T_{min} only where statistically relevant. The cluster analysis reveals that the majority of the stations, 23 out of the 35, are most appropriately classified into four seasons (Table 3). The remaining 12 stations are best classified into three seasons. All the stations in the Limpopo, KwaZulu-Natal and North-West Provinces are clustered into four seasons, and those in the Eastern Cape are clustered into three seasons. All these stations have a statistically strong grouping (CPCC>0.7) and distinct cluster structures (ASW>0.5), except for Cedara (ASW=0.47) in KwaZulu-Natal, with a weaker cluster structure. The weaker cluster structures are also prominent in the Eastern Cape stations (ASW<0.5). The cluster analysis results revealed that Dohne (CPCC=0.7009) in the Eastern Cape has the lowest quality of grouping among the 35 stations analysed.

The three stations in the Free State, two of which (Bethlehem and Welkom) are classified into four seasons and one (Bloemfontein) into three seasons, have a good quality grouping (CPCC>0.7) and distinct cluster structure (ASW>0.5). A similar degree of confidence in cluster structures is found in both the Mpumalanga stations, Skukuza and Carolina, which are divided into three and four seasons, respectively. These two stations have a higher quality grouping (CPCC>0.8) than those in the Free State. A higher quality grouping is also prevalent in both the Gauteng stations, with Johannesburg Int (International) classified into four seasons, and Zuurbekom into three. However, a weaker cluster structure is calculated for Zuurbekom (ASW=0.45).

The Northern Cape and Western Cape are noticeably different from the rest of the provinces, with a more significant variation in the degree of confidence for the cluster structures. Six of the nine stations in the Northern Cape are clustered into four seasons with good cluster groupings (CPCC>0.7). However, most of these stations have a weak cluster structure (ASW<0.5), except for Fraserburg (ASW=0.5). The remaining three stations – De Aar, Kimberley and Springbok – are divided into three seasons with a good quality grouping; however, weaker structures are visible in Kimberley (ASW=0.46) and Springbok (ASW=0.42). In the Western Cape, four out of the six stations are classified into four seasons and the other two (Mosselbay and Cape Columbine) into three seasons. Similar to the Northern Cape, all stations clustered into four seasons display a good quality of grouping, but a weak cluster structure (ASW<0.5), with the weakest cluster structure calculated for Cape Agulhas (ASW=0.46). Mosselbay (ASW=0.51; CPCC=0.8011) is the only station in the WC with a distinct cluster structure and a good quality grouping. By contrast, the cluster analysis identified Cape Columbine (ASW=0.4) with the weakest cluster structure among all the stations that were analysed.

While similarities are found in the number of seasonal groupings returned for the different temperature metrics, the highest number (15 stations) have consistent classifications for T_{min} and T_{avg}. Only six stations have consistent seasonal classifications when considering T_{max} and $T_{min,}$ and three stations when considering T_{max} and T_{avg}. Only seven stations – Dohne, Port Elizabeth, De Aar, Bloemfontein, Cape Columbine, Welkom and Escourt – have the same number of seasonal groupings for all three temperature metrics.

For the majority of stations (23), classification using T_{max} returns three seasons, with 9 stations classified into two seasons, and only 3 classified into four seasons (Table 3). The majority of these stations have a strong grouping (CPCC>0.7), except for Dohne (CPCC=0.6714), located in the Eastern Cape. However, the degree of confidence in the cluster structures is low due to the weak cluster structures (ASW<0.5) for most of the stations except for Johannesburg Int, Zuurbekom, Mara, Mahikeng, Springbok and Cape Town with an ASW>0.5. The cluster analysis for T_{min} classifies the majority of the stations (21) into four seasons, with 13 stations classified into three seasons, and only 1 station (Springbok) classified into two seasons. Similar to T_{max}, the grouping quality for the stations is good. However, the cluster structures for most of the stations are distinct with only 19 stations returning an ASW<0.5.

Spatial analysis of the cluster analysis results (Figure 2) indicates that most parts of the country experience three seasons, with the greatest spatial variability visible in T_{max}. Similarities in the classification of seasons are visible for T_{max}, T_{min} and T_{avg}, but more so for T_{min} and T_{avg}. The western and central regions of the country, and parts of the Eastern Cape, have three distinct seasons, when classified using T_{max}, T_{min} and T_{avg}. Areas surrounding Springbok are similarly classified as having only two distinct seasons.

Table 3: Output of the cluster analysis including cluster groups (CG), average silhouette widths (ASW) and cophenetic correlation coefficient (CPCC) scores

Stations		T_{max}				T_{min}				T_{avg}			
		CG	ASW	CPCC	Seasonal Groups	CG	ASW	CPCC	Seasonal Groups	CG	ASW	CPCC	Seasonal Groups
EASTERN CAPE	East London	2	0.47	0.7293	May/Jun/Jul/Aug/Sept/Oct/Nov Dec/Jan/Feb/Mar/Apr	3	0.5	0.7487	May/Jun/Jul/Aug/Sep Apr/Oct/Nov Dec/Jan/Feb/Mar	3	0.45	0.7475	Dec/Jan/Feb/Mar Jun/Jul/Aug Apr/May/Sep/Oct/Nov
	Dohne	3	0.37	0.6714	Dec/Jan/Feb/Mar Jun/Jul/Aug Apr/May/Sep/Oct/Nov	3	0.42	0.7091	Dec/Jan/Feb/Mar Jun/Jul/Aug Apr/May/Sept/Oct/Nov	3	0.44	0.7009	Dec/Jan/Feb/Mar Jun/Jul/Aug Apr/May/Sept/Oct/Nov
	Port Elizabeth	3	0.38	0.7833	Dec/Jan/Feb/Mar/Apr Jun/Jul/Aug/Sep May/Oct/Nov	3	0.49	0.7627	May/Jun/Jul/Aug/Sep Apr/Oct/Nov Dec/Jan/Feb/Mar	3	0.47	0.7385	Dec/Jan/Feb/Mar Jun/Jul/Aug Apr/May/Sep/Oct/Nov
	Mthatha	2	0.39	0.6644	Dec/Jan/Feb/Mar/Apr/Nov May/Jun/Jul/Aug/Sep/Oct	4	0.48	0.7663	Jun/Jul May/Aug/Sep Apr/Oct/Nov Dec/Jan/Feb/Mar	3	0.42	0.7125	May/Jun/Jul/Aug Dec/Jan/Feb/Mar Apr/Sep/Oct/Nov
FREE STATE	Bethlehem	3	0.34	0.8042	May/Jun/Jul/Aug Dec/Jan/Feb Mar/Apr/Sep/Oct/Nov	4	0.48	0.8057	Jun/Jul May/ Aug Apr/Sep/Oct Dec/Jan/Feb/Mar/Nov	4	0.53	0.8149	Jun/Jul May/Aug Apr/Sep Dec/Jan/Feb/Mar/Oct/Nov
	Bloemfontein	3	0.34	0.7867	May/Jun/Jul/Aug Dec/Jan/Feb Mar/Apr/Sep/Oct/Nov	3	0.48	0.7565	Dec/Jan/Feb/Mar/Oct/Nov Apr/Sep May/Jun/Jul/Aug	3	0.5	0.7545	Dec/Jan/Feb/Mar/Oct/Nov Apr/Sep May/Jun/Jul/Aug
	Welkom	4	0.44	0.7066	Dec/Jan/Feb/Mar/Oct/Nov Jun/Jul May/Aug Apr/Sep	4	0.48	0.7415	Dec/Jan/Feb/Mar/Oct/Nov Apr/Sep Jun/Jul May/Aug	4	0.54	0.8245	Jun/Jul May/Aug Apr/Sep Dec/Jan/Feb/Mar/Oct/Nov

Table 3 continues on the next page

Table 3 continued

Stations					T_{max}				T_{min}				T_{avg}
		CG	ASW	CPCC	Seasonal Groups	CG	ASW	CPCC	Seasonal Groups	CG	ASW	CPCC	Seasonal Groups
GAUTENG	Johannesburg Int	3	0.5	0.8364	Dec/Jan/Feb/Mar/Apr/Sep/Oct/Nov Jun/Jul May/Aug	3	0.45	0.8170	May/Jun/Jul/Aug Apr/Sep Dec/Jan/Feb/Mar/Oct/Nov	4	0.51	0.8458	Jun/Jul May/Aug Apr/Sep Dec/Jan/Feb/Mar/Oct/Nov
	Zuurbekom	2	0.53	0.8299	May/Jun/Jul/Aug Dec/Jan/Feb/Mar/Apr/Sep/Oct/Nov	4	0.47	0.8307	May/Jun/Jul/Aug Apr/Sep Dec/Jan/Feb Mar/Oct/Nov	3	0.45	0.8357	May/Jun/Jul/Aug Dec/Jan/Feb/Mar/Oct/Nov Apr/Sep
KWAZULU-NATAL	Cedara	2	0.37	0.6561	Dec/Jan/Feb/Mar Apr/May/Jun/Jul/Aug/Sep/Oct/Nov	4	0.52	0.8056	Jun/Jul May/Aug Apr/Sept/Oct Dec/Jan/Feb/Mar/Nov	4	0.47	0.7686	Jun/Jul May/ Aug Dec/Jan/Feb/Mar Apr/Sep/Oct/Nov
	Durban	3	0.47	0.7316	Jun/Jul/Aug/Sep/Oct Dec/Jan/Feb/Mar Apr/May/Nov	4	0.52	0.7680	Jun/Jul May/Aug/Sep Apr/Oct/Nov Dec/Jan/Feb/Mar	4	0.52	0.7125	Dec/Jan/Feb/Mar Jun/Jul/Aug Apr/Nov May/Oct/Sep
	Escourt	4	0.37	0.8278	Jun/Jul May/Aug Apr/Sep/Oct Dec/Jan/Feb/Mar/Nov	4	0.54	0.8278	May/Jun/Jul/Aug Apr/Sep Dec/Jan/Feb Mar/Oct/Nov	4	0.54	0.8212	Jun/Jul May/Aug Apr/Sept/Oct Dec/Jan/Feb/Mar/Nov
	Shaleburn	2	0.46	0.7406	May/Jun/Jul/Aug Dec/Jan/Feb/Mar/Apr/Sep/Oct/Nov	4	0.42	0.8214	May/Jun/Jul/Aug Apr/Sept Dec/Jan/Feb Mar/Oct/Nov	4	0.52	0.8103	Jun/Jul May/Aug Apr/Sep/Oct Dec/Jan/Feb/Mar/Nov
LIMPOPO	Mara	2	0.52	0.8130	Dec/Jan/Feb/Mar/Apr/Sep/Oct/Nov May/Jun/Jul/Aug	4	0.50	0.8444	Jun/Jul May/ Aug Dec/Jan/Feb/Mar/Nov Apr/Sep/Oct	4	0.5	0.8460	Jun/Jul May/ Aug Dec/Jan/Feb/Mar/Nov Apr/Oct/Sep
	Polokwane	3	0.45	0.8186	Dec/Jan/Feb/Mar/Apr/Sep/Oct/Nov Jun/Jul May/Aug	3	0.55	0.8248	Dec/Jan/Feb/Mar/Nov Apr/Sep/Oct May/Jun/Jul/Aug	4	0.5	0.8332	Jun/Jul May/Aug Apr/Sep Dec/Jan/Feb/Mar/Oct/Nov
	Warmbad Towoomba	3	0.48	0.7906	Dec/Jan/Feb/Mar/Sep/Oct/Nov Jun/Jul Apr/May/Aug	4	0.56	0.8526	May/Aug Jun/Jul Apr/Sep Dec/Jan/Feb/Mar/Oct/Nov	4	0.57	0.8729	Jun/Jul May/Aug Apr/Sep Dec/Jan/Feb/Mar/Oct/Nov
MPUMALANGA	Skukuza	2	0.48	0.7837	May/Jun/Jul/Aug Dec/Jan/Feb/Mar/Apr/Sep/Oct/Nov	4	0.51	0.8365	Jun/Jul May/Aug Dec/Jan/Feb/Mar/Nov Apr/Sep/Oct	3	0.51	0.8307	May/Jun/Jul/Aug Apr/Sep/Oct Dec/Jan/Feb/Mar/Nov
	Carolina	3	0.46	0.7622	Dec/Jan/Feb/Mar/Sep/Oct/Nov Jun/Jul Apr/May/Aug	4	0.52	0.8357	May/Aug Jun/Jul Apr/Sep/Oct Dec/Jan/Feb/Mar/Nov	4	0.5	0.8440	Jun/Jul May/Aug Apr/Sep Dec/Jan/Feb/Mar/Oct/Nov
NORTH-WEST	Mahikeng	2	0.57	0.8271	May/Jun/Jul/Aug Dec/Jan/Feb/Mar/Apr/Sep/Oct/Nov	4	0.56	0.8424	Jun/Jul May/Aug Apr/Sep Dec/Jan/Feb/Mar/Oct/Nov	4	0.58	0.8514	Jun/Jul May/Aug Apr/Sep Dec/Jan/Feb/Mar/Oct/Nov
	Vryburg	4	0.46	0.8226	Jun/Jul May/Aug Apr/Sep Dec/Jan/Feb/Mar/Oct/Nov	3	0.5	0.7625	Dec/Jan/Feb/Mar/Oct/Nov Apr/Sep May/Jun/Jul/Aug	4	0.58	0.8348	Jun/Jul May/Aug Apr/Sep Dec/Jan/Feb/Mar/Oct/Nov

Table 3 continues on the next page

Table 3 continued

Stations		T_{max}				T_{min}				T_{avg}			
		CG	ASW	CPCC	Seasonal Groups	CG	ASW	CPCC	Seasonal Groups	CG	ASW	CPCC	Seasonal Groups
NORTHEN CAPE	Calvinia	3	0.47	0.7246	Dec/Jan/Feb/Mar/Nov Jun/Jul/Aug Apr/May/Oct/Sep	3	0.47	0.7567	Dec/Jan/Feb/Mar Jun/Jul/Aug/Sep Apr/May/Oct/Nov	4	0.48	0.7864	Jun/Jul/Aug May/Sep Apr/Oct/Nov Dec/Jan/Feb/Mar
	De Aar	3	0.44	0.7561	Dec/Jan/Feb/Mar/Oct/Nov Apr/Sep May/Jun/Jul/Aug	3	0.49	0.7198	Dec/Jan/Feb/Mar/Nov Jun/Jul/Aug Apr/May/Oct/Sep	3	0.5	0.7209	Dec/Jan/Feb/Mar/Nov Apr/Sep/Oct May/Jun/Jul/Aug
	Fraserburg	3	0.48	0.7297	Dec/Jan/Feb/Mar/Nov Apr/Sep/Oct May/Jun/Jul/Aug	4	0.53	0.8077	Jun/Jul/Aug May/Sep Apr/Oct/Nov Dec/Jan/Feb/Mar	4	0.5	0.7349	Dec/Jan/Feb/Mar/Nov Jun/Jul/Aug Apr/Oct May/Sept
	Kimberley	3	0.41	0.7382	Dec/Jan/Feb/Mar/Oct/Nov Apr/Sep May/Jun/Jul/Aug	4	0.46	0.7703	Jun/Jul May/Aug Apr/Sep/Oct Dec/Jan/Feb/Mar/Nov	3	0.46	0.7536	Dec/Jan/Feb/Mar/Oct/Nov Apr/Sep May/Jun/Jul/Aug
	Pofadder	3	0.43	0.7750	Jun/Jul/Aug/Sep Dec/Jan/Feb/Mar/Nov Apr/Oct	4	0.48	0.7947	Jun/Jul/Aug May/Sep Apr/Oct/Nov Dec/Jan/Feb/Mar	4	0.48	0.7967	May/Sep Jun/Jul/Aug Dec/Jan/Feb/Mar Apr/Oct/Nov
	Springbok	2	0.55	0.7802	May/Jun/Jul/Aug/Sep Dec/Jan/Feb/Mar/Apr/Oct/Nov	2	0.49	0.7412	Dec/Jan/Feb/Mar/Apr May/Jun/Jul/Aug/Sep/ Oct/Nov	3	0.42	0.7426	Dec/Jan/Feb/Mar/Apr/Nov Jun/Jul/Aug May/Sep/Oct
	Sutherland	3	0.46	0.7292	Dec/Jan/Feb/Mar/Nov Apr/May/Sep/Oct Jun/Jul/Aug	3	0.51	0.7618	May/Jun/Jul/Aug/Sep Apr/Nov/Oct Dec/Jan/Feb/Mar	4	0.49	0.7774	Jun/Jul/Aug May/Sep Apr/Oct Dec/Jan/Feb/Mar/Nov
	Upington	3	0.42	0.7494	Dec/Jan/Feb/Mar/Oct/Nov Apr/Sep May/Jun/Jul/Aug	4	0.51	0.8192	May/Sep Jun/Jul/Aug Dec/Jan/Feb/Mar Apr/Oct/Nov	4	0.46	0.7155	Dec/Jan/Feb/Mar/Nov Jun/Jul/Aug May/Sep Apr/Oct
	Vanwyksvlei	3	0.48	0.7221	Dec/Jan/Feb/Mar/Nov Apr/Sep/Oct May/Jun/Jul/Aug	4	0.51	0.7300	Dec/Jan/Feb/Mar/Nov Jun/Jul/Aug Apr/Oct May/Sep	4	0.47	0.7276	Dec/Jan/Feb/Mar/Nov Jun/Jul/Aug Apr/Oct May/Sep
WESTERN CAPE	Cape Agulhas	3	0.43	0.7670	May/Jun/Jul/Aug/Sep/Oct Dec/Jan/Feb/Mar Apr/ Nov	4	0.48	0.7685	Dec/Jan/Feb/Mar Jun/Jul/Aug/Sep Apr/Nov May/Oct	4	0.46	0.7741	Dec/Jan/Feb/Mar Jun/Jul/Aug/Sep Apr/Nov May/Oct
	Cape Columbine	3	0.41	0.7861	Dec/Jan/Feb/Mar/Apr/Nov Jun/Jul/Aug May/ Sep/Oct	3	0.37	0.7612	Dec/Jan/Feb/Mar/Apr/Nov May/Oct Jun/Jul/Aug/Sep	3	0.4	0.7666	Jun/Jul/Aug/Sep Dec/Jan/Feb/Mar Apr/May/Oct/Nov
	Cape Town	3	0.53	0.7893	May/Jun/Jul/Aug/Sep Apr/Oct/Nov Dec/Jan/Feb/Mar	3	0.5	0.7516	May/Jun/Jul/Aug/Sep Dec/Jan/Feb/Mar Apr/Oct/Nov	4	0.49	0.7757	Jun/Jul/Aug May/ Sep Apr/Oct/Nov Dec/Jan/Feb/Mar
	George	3	0.42	0.7453	Dec/Jan/Feb/Mar Apr/May/Nov Jun/Jul/Aug/Sep/Oct	4	0.48	0.7606	Dec/Jan/Feb/Mar Jun/Jul/Aug/Sep Apr/Nov May/Oct	4	0.49	0.7635	Dec/Jan/Feb/Mar Jun/Jul/Aug/Sep Apr/Nov May/Oct
	Mosselbay	3	0.42	0.8017	Dec/Jan/Feb/Mar Jun/Jul/Aug/Sep/Oct Apr/May/Nov	4	0.47	0.7862	Dec/Jan/Feb/Mar Jun/Jul/Aug/Sep May/Oct Apr/ Nov	3	0.51	0.8002	Dec/Jan/Feb/Mar Jun/Jul/Aug/Sep Apr/May/Oct/Nov
	Vredendal	3	0.44	0.7630	Dec/Jan/Feb/Mar/Apr/Nov Jun/Jul/Aug May/Sep/Oct	3	0.47	0.7862	Dec/Jan/Feb/Mar May/ Jun/Jul/Aug/Sep Apr/Oct/Nov	4	0.47	0.8011	Dec/Jan/Feb/Mar Jun/Jul/Aug May/Sep Apr/Oct/Nov

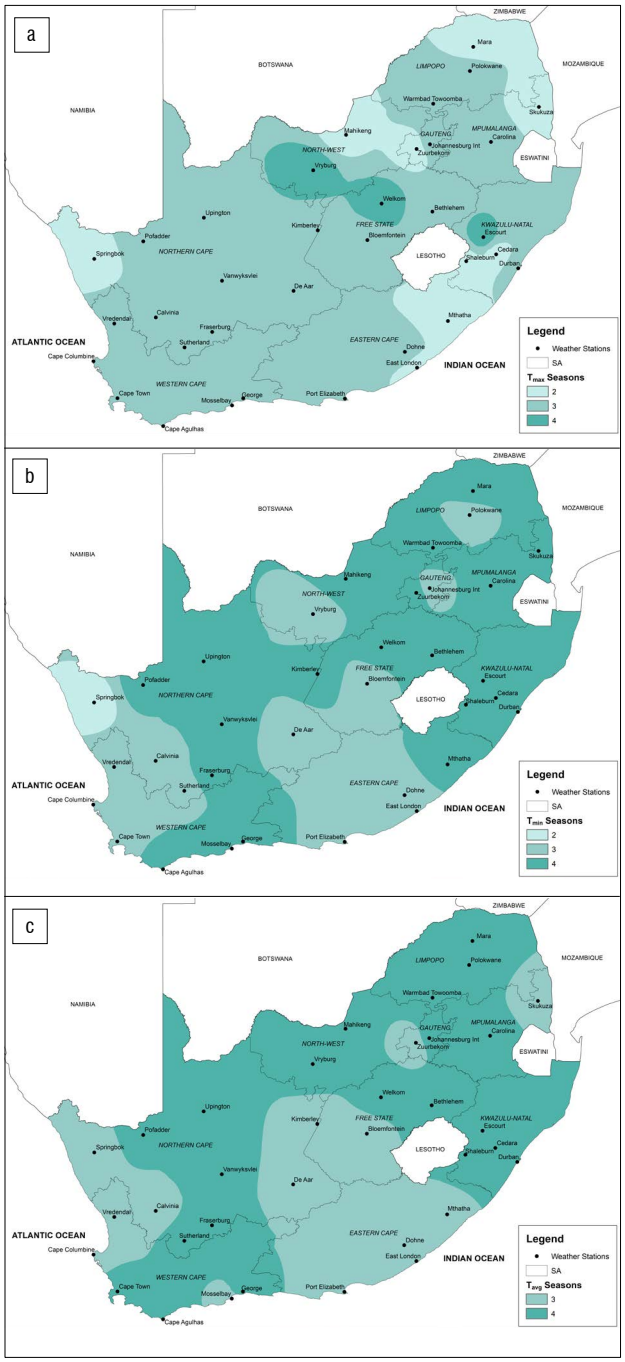


Figure 2: (a) Maximum (T_{max}), (b) minimum (T_{min}) and (c) average (T_{avg}) temperature seasonal groupings.

Seasonal timetable

Several variations of monthly classifications have been calculated (Tables 3 and 4), which will be referred to as 'seasonal brackets'.

Summer: There is unanimous agreement across seasons and temperature metrics that December, January, February and March can be grouped into the summer seasonal bracket. October and November can tentatively be included in the summer season, with a 31% agreement among stations for October and an overall 59% agreement for November. A total of 30% of stations classify October as falling within spring, 26% in late spring, 5% in early spring, and 8% in the winter seasonal bracket. In comparison, 21% of the remaining stations classify November as falling in spring, 17% in late spring, and 3% in winter. The latest start of summer is December. The longest summer season is calculated for Mthatha (Eastern Cape) using T_{avg} as the classifier, spanning the earliest start month of September and persisting until April.

Autumn: For 43% of the stations, April falls within the early autumn bracket, whereas for 39% of stations, April is grouped in the main autumn seasonal bracket. The remainder of the stations group April into summer (15%), winter (2%) and late autumn (1%). A total of 41% of the stations classify May into the late autumn bracket, 35% into the winter bracket, and 24% into the main autumn bracket. Autumn starts as early as March for Bethlehem, Bloemfontein and Zuurbekom using T_{max} and T_{min} as the classifier. Using T_{avg} as the classifier, the majority of stations indicate the start of autumn in April. Springbok (Northern Cape) has the latest classified start of autumn, in the month May. Stations with only 1 month of autumn, using T_{avg} as classifier, include Bloemfontein, Zuurbekom, Skukuza, De Aar, Kimberley and Springbok

Winter: There is unanimous agreement across stations and temperature metrics that June and July are grouped into the winter bracket. There is 71% agreement among stations that August additionally forms part of the winter bracket with only 25% of stations classifying August in early spring and 4% in the main spring seasonal bracket. Cedara and Vanwyksvlei experience the longest winter, commencing in April and ending in November for Cedara and October for Vanwyksvlei, both using T_{max} as the classifier. Winter in East London and Springbok starts a month later in May and ends in November. Stations with the shortest winter (June to July) classified by T_{avg} include Shaleburn, Mara, Polokwane, Warmbad Towoomba, Carolina, Bethlehem, Welkom, Johannesburg Int, Cedara, Escourt, Mahikeng, Vryburg, Cape Agulhas, Cape Columbine and George.

Spring: Interestingly, the most disputed month across all stations is September. However, 29% of the stations agree that September can tentatively be included in the spring bracket. The remainder comprise a 25% agreement that September can be included in the winter, 23% in early spring, 16% in late spring and 9% in the summer bracket. Some stations are classified as experiencing early spring in August, using T_{min} as the classifier. These stations are Mthatha, Bethlehem, Welkom, Cedara, Durban, Escourt, Mara, Warmbad Towoomba, Skukuza, Carolina, Mahikeng and Kimberley. Stations with the shortest spring (only the month of August), calculated using T_{max}, are Johannesburg Int, Polokwane, Warmbad Towoomba, and Carolina. Interestingly, all these stations are situated in the northern parts of the country. Classified by T_{min}, the longest spring is observed in Mthatha (Eastern Cape) and Durban (KwaZulu-Natal), starting in August and ending in November. The latest start for spring – November – classified by T_{max}, is visible in Durban, Cape Agulhas and Mosselbay. Stations at East London, Mthatha, Zuurbekom, Cedara, Shaleburn, Mara, Skukuza, Mahikeng and Springbok each have two classified seasons, and therefore no autumn or spring.

Start and end dates of summer and winter

A distinct, southwest to northeast spatial pattern is apparent for the start and end dates of summer and winter across all the temperature metrics (Figures 3a–f and 4a–f). Summer broadly commences earlier in the northeastern and interior parts of the country and later along the southwestern parts and the south coast. The earliest start of summer is visible in T_{max} for the northern parts of the country, and in parts of KwaZulu-Natal for T_{avg}. The earliest end of summer is calculated using T_{min}, with parts in KwaZulu-Natal and Gauteng ending in February (Figure 3e). The greatest variability in the spatial patterns is recorded T_{max}, for which the northern and southern region summer ends in April, similar to T_{avg} for the southern region. However, there is a consensus amongst the temperature metrics that, for most parts of the country, summer ends in March, whereas for the western parts of the country, the season ends a month later in April.

For the majority of the country, winter starts in June, with the season starting earlier for a few interior regions. The greatest spatial variability in

the timing of the start of winter is observed in T_{max} (Figure 4a). For areas in the Western Cape and Northern Cape, winter is classified as starting in May, similar to the start of winter using T_{min}. For parts of KwaZulu-Natal, winter is calculated to start as early as April. Regarding the end of winter, the greatest spatial variability is similarly observed for calculations using T_{max}, for which winter ends latest in the southwest of the country and along the east coast. For the western parts of the Northern Cape, winter ends later using T_{max} and T_{min}. Similar to the end of summer maps (Figure 4d–f), a distinct southwest to northeast spatial movement of end dates is visible for all the temperature metrics used. For the southwestern parts of the country, winter ends later, whereas moving northeastwards to the interior, the winter months end earlier, except for some parts in Gauteng, Limpopo and the North-West.

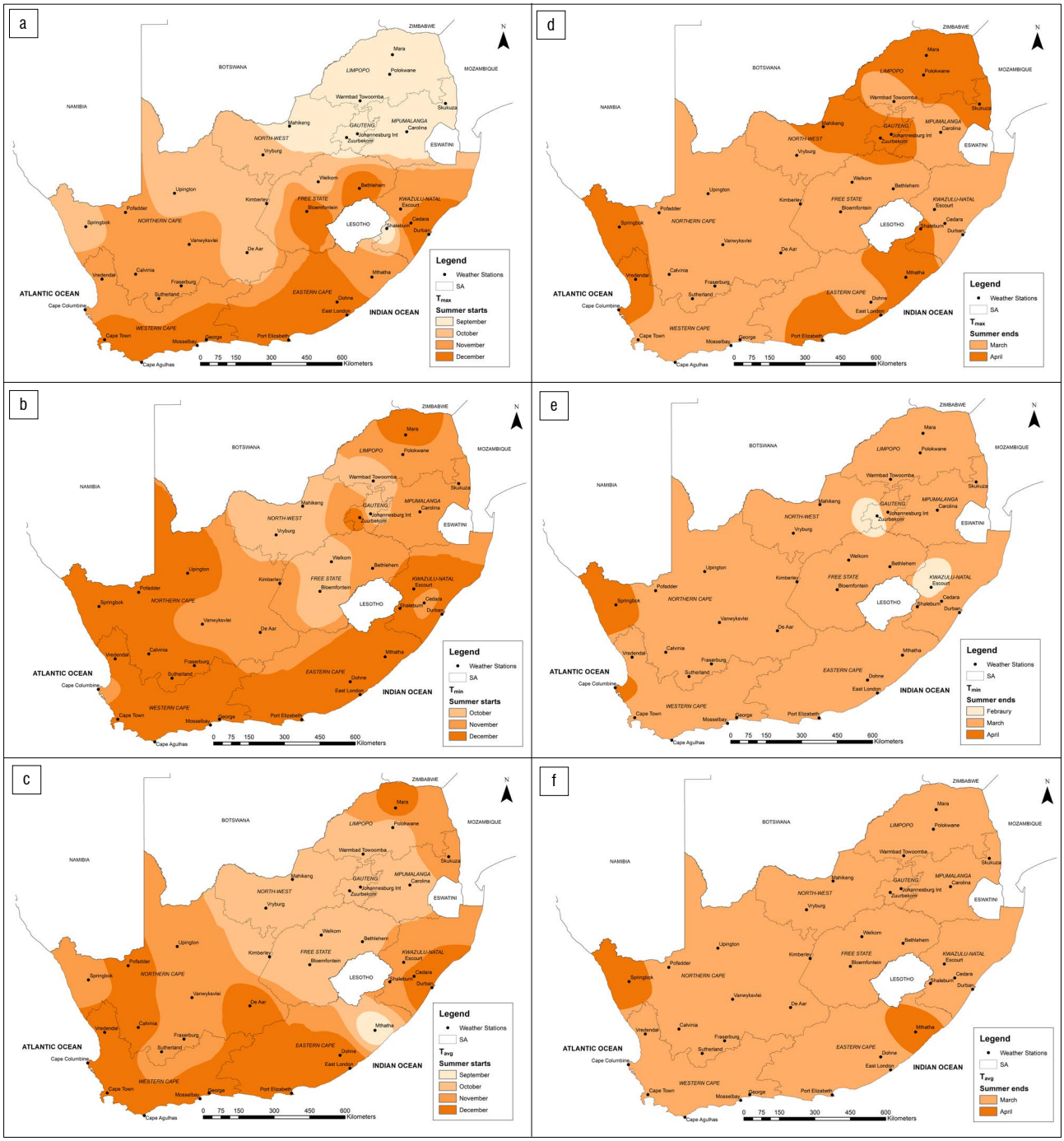


Figure 3: (a–c) Start of summer and (d–f) end of summer.

Table 4: Seasonal timetable of maximum (T_{max}), minimum (T_{min}) and average (T_{avg}) temperatures

	Station		Jan	Feb	Mar	Apr	May	June	July	Aug	Sep	Oct	Nov	Dec
EASTERN CAPE	East London	T_{max}												
		T_{avg}												
		T_{min}												
	Dohne	T_{max}												
		T_{avg}												
		T_{min}												
	Port Elizabeth	T_{max}												
		T_{avg}												
		T_{min}												
	Mthatha	T_{max}												
		T_{avg}												
		T_{min}												
FREE STATE	Bethlehem	T_{max}												
		T_{avg}												
		T_{min}												
	Bloemfontein	T_{max}												
		T_{avg}												
		T_{min}												
	Welkom	T_{max}												
		T_{avg}												
		T_{min}												
GAUTENG	Johannesburg Int	T_{max}												
		T_{avg}												
		T_{min}												
	Zuurbekom	T_{max}												
		T_{avg}												
		T_{min}												
KWAZULU-NATAL	Cedara	T_{max}												
		T_{avg}												
		T_{min}												
	Durban	T_{max}												
		T_{avg}												
		T_{min}												
	Escourt	T_{max}												
		T_{avg}												
		T_{min}												
	Shaleburn	T_{max}												
		T_{avg}												
		T_{min}												
LIMPOPO	Mara	T_{max}												
		T_{avg}												
		T_{min}												
	Polokwane	T_{max}												
		T_{avg}												
		T_{min}												
	Warmbad Towoomba	T_{max}												
		T_{avg}												
		T_{min}												
MPUMALANGA	Skukuza	T_{max}												
		T_{avg}												
		T_{min}												
	Carolina	T_{max}												
		T_{avg}												
		T_{min}												
NORTH-WEST	Mahikeng	T_{max}												
		T_{avg}												
		T_{min}												
	Vryburg	T_{max}												
		T_{avg}												
		T_{min}												
NORTHERN CAPE	Calvinia	T_{max}												
		T_{avg}												
		T_{min}												
	De Aar	T_{max}												
		T_{avg}												
		T_{min}												
	Fraserburg	T_{max}												
		T_{avg}												
		T_{min}												
	Kimberley	T_{max}												
		T_{avg}												
		T_{min}												
	Pofadder	T_{max}												
		T_{avg}												
		T_{min}												
	Springbok	T_{max}												
		T_{avg}												
		T_{min}												
	Sutherland	T_{max}												
		T_{avg}												
		T_{min}												
	Upington	T_{max}												
		T_{avg}												
		T_{min}												
	Vanwyksvlei	T_{max}												
		T_{avg}												
		T_{min}												
WESTERN CAPE	Cape Agulhas	T_{max}												
		T_{avg}												
		T_{min}												
	Cape Columbine	T_{max}												
		T_{avg}												
		T_{min}												
	Cape Town	T_{max}												
		T_{avg}												
		T_{min}												
	George	T_{max}												
		T_{avg}												
		T_{min}												
	Mosselbay	T_{max}												
		T_{avg}												
		T_{min}												
	Vredendal	T_{max}												
		T_{avg}												
		T_{min}												

Key

Summer	Autumn	Late Spring	
Early Autumn	Winter	Spring	
Late Autumn	Early Spring		

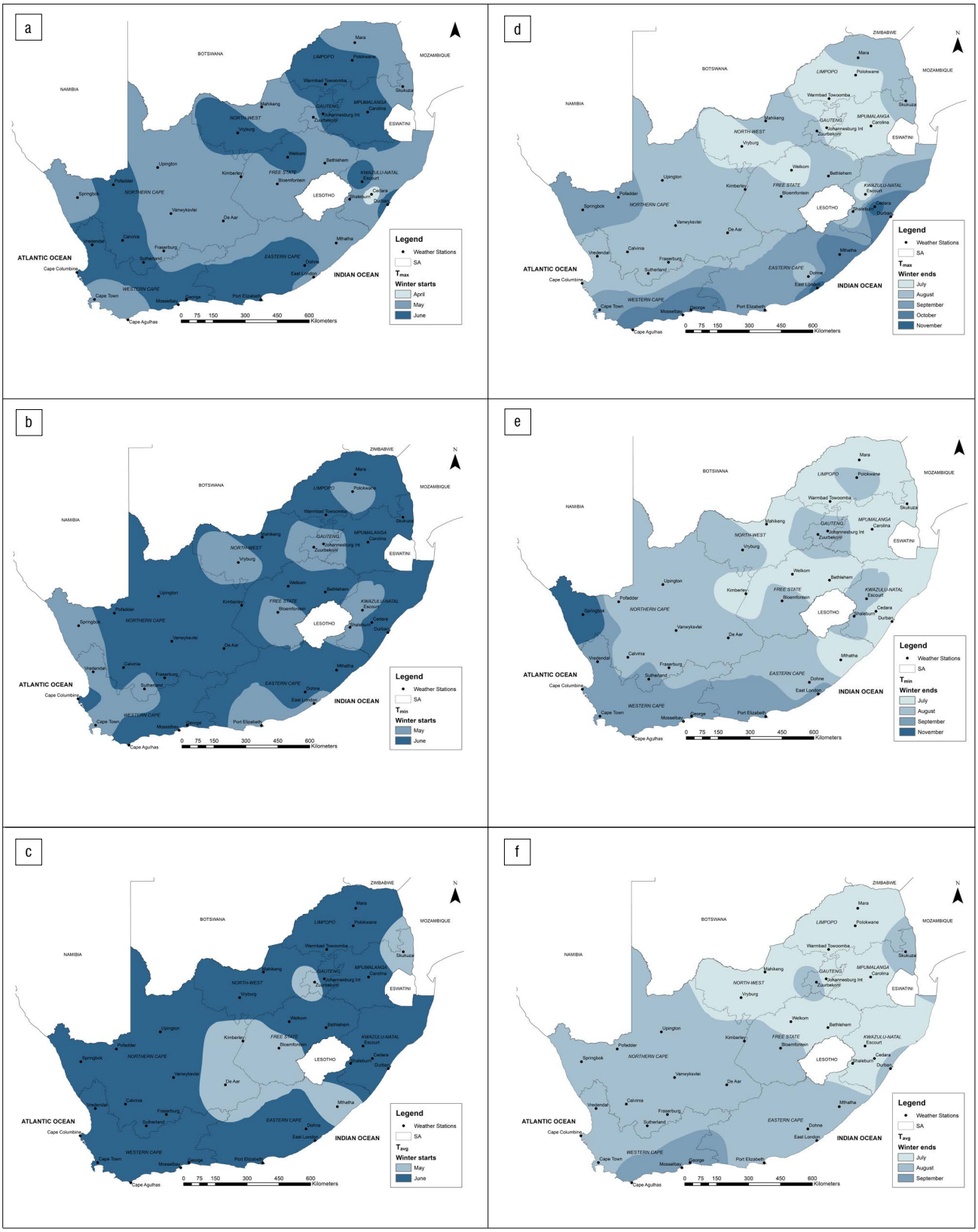


Figure 4: (a–c) Start of winter and (d–f) end of winter.

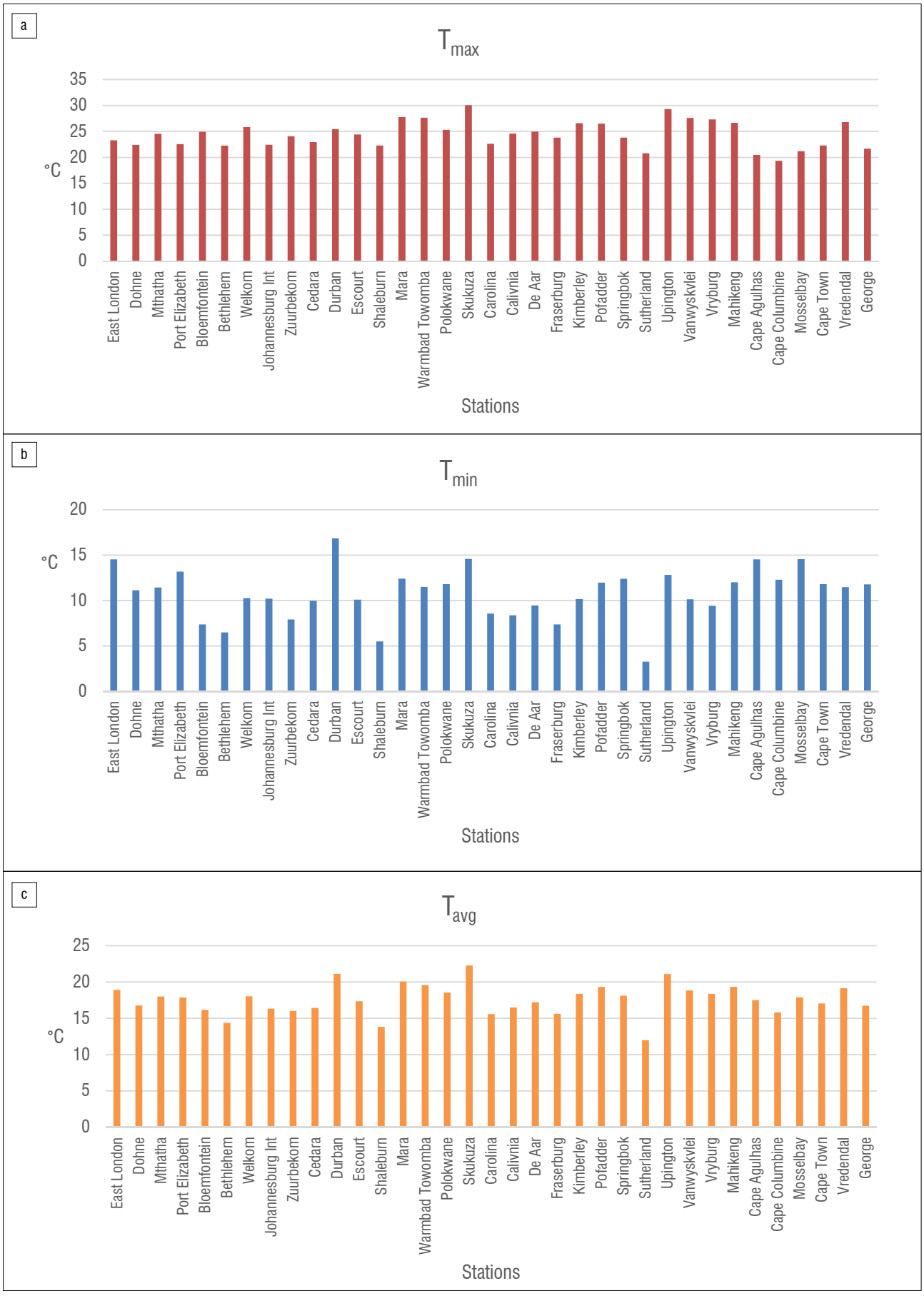


Figure 5: Annual mean of (a) maximum (T_{max}), (b) minimum (T_{min}) and (c) average (T_{avg}) temperatures for the selected stations, 1980–2015.

Discussion and conclusion

We present one of the first statistical classifications of seasons across South Africa using daily temperatures. Daily temperature data across the country were used as a distinctive marker to classify the seasons due to the detectability of temperature changes compared to rainfall across South Africa. Through statistical analysis and results captured in the seasonal timetable (Table 4), new seasonal brackets are put forward in accordance with the agreement of seasons and temperatures among stations used in this research.

Aggregated for the whole country, based on T_{max}, T_{min} and T_{avg}, our results show that the weather stations agree that the following seasonal brackets can be used:

- Summer (October/November/December/January/February/March)
- Early autumn (April)
- Late autumn (May)
- Winter (June/July/August)
- Spring (September)

These proposed seasonal brackets challenge our 'common knowledge' of four equal length seasons of 3 months each[2,26-29], and the ad-hoc approaches some researchers use in South Africa[25-27]. Noticeable similarities occur between the two seasonal divisions of months used to define farming seasons[30] as well as monthly summer divisions related to the positions of South Africa related to disease-risk seasons.[7] However, the proposed longer duration of summer and shorter spring seasons may conflict with the agricultural practices used currently, in particular, the current observed length and timing of the growing season across the country.[1-3] Additionally, these proposed seasonal brackets may assist in the explanation of current delays and advances in seasonal phenological events[33], and challenges in the tourism sector where most outdoor attractions are dependent on the seasonal climate[5].

However, the high spatio-temporal variability in temperatures (e.g. annual mean temperatures Figure 5) presents a complex picture of seasonality. This presents challenges in defining seasonal brackets for a given location or region, particularly where regional climate regimes change within a small geographic area[24], and due to the complexity of South Africa's climate[29]. Discrepancies have been found among the different temperature metrics. However, the majority of the stations (23 out of the 35), are divided into four seasons, using T_{avg} as the classifier, with the remaining 12 stations clustered into three seasons. Interestingly, some stations within the same province (e.g. Johannesburg Int and Zuurbekom in Gauteng) have different seasonal groupings. With closer inspection, these differences may occur due to the location and elevation of the stations (Table 2). For example, it has been found that built-up areas such as Johannesburg may be warmer in late winter than rural areas due to the urban heat island[52] and higher elevations tend to be cooler than lower elevations[53]. Taking the above-mentioned into consideration, the importance of selecting the relevant temperature metric, e.g. T_{max}, T_{min} and T_{avg}, is highlighted for analysis purposes, as this selection can return different results as portrayed in the results.

In general, the findings of the start and end dates of summer and winter (Figures 3 and 4) coincide with the pressure regimes, as well as the interannual migration of the ITCZ.[19,37] The results indicate that summer starts later (ending earlier) and winter starts earlier (ending later) in the southwestern and southern regions of the country. These results coincide with the movement of the cold front of the mid-latitude cyclones during the winter months.[38] While, during summer, the southward movement of the ITCZ and the position of the subtropical high-pressure system are associated with warmer conditions, which may result in the patterns found. Summers start earlier, and winters start and end later in the northeastern parts of the county. These patterns are found independently from the notable link between temperatures and weather systems. The patterns also show the annual progression of temperatures which follow a southwest to a northeastwards spatial pattern across the country.

The key limitations of this study are the nature of the temperature data sets. The data sets are not perfect and inherent errors may be present for a number of reasons.[29] Furthermore, inhomogeneity is not likely to play a significant role in this study as the consistency was ensured by using only SAWS data sets.[54] Mean daily temperature data were quantified using T_{max} and T_{min}; this is a limitation as hourly temperature readings may provide accurate values of mean daily temperatures.[54] Furthermore, we acknowledge that station measurements are unable to display complete areal coverage as these are location-specific[54,55], which is particularly an issue for the interpolated maps presented throughout. A limited number of stations that have long-term temperature records was selected using a broad grid approach, as discussed, to get a relatively good spatial representation of the country. To overcome this limitation, future research may benefit from the inclusion of temperature data from additional weather stations from other organisations, such as the South African Agricultural Research Council. Such addition would, however, require greater efforts at data homogenisation and quality checking, which introduce a further set of limitations.

Finally, this research provides an insight into the complexity of seasonality across South Africa, as well as direction for climate-relevant research with temperature data as the primary input. Possibly the most significant contribution of this research is the newly proposed seasonal brackets using temperature metrics. The knowledge presented here is crucial for agriculture practices, resource management, tourism and other temperature-dependent activities, especially to develop adaptive strategies in monitoring seasonal changes in temperatures under climate change.

Acknowledgements

We acknowledge the financial and collegial support offered to A.v.d.W. from the Faculty of Natural and Agricultural Sciences and the Geography Department at the University of the Free State. J.M.F. received funding from the DSI-NRF Centre of Excellence for Palaeoscience. We thank Professor Christopher Curtis for advice on earlier stages of the project.

Competing interests

We declare that there are no competing interests.

Authors' contributions

A.v.d.W.: Data collection, data analysis, data curation, validation, writing – the initial draft, writing – revisions. J.M.F.: Conceptualisation, methodology, validation, writing – revisions, student supervision, project leadership.

Data availability

Data are owned by the South African Weather Service and can be obtained from them on request.

References

1. Battisti DS, Naylor RL. Historical warmings of future food insecurity with unprecedented seasonal heat. Science. 2009;323(5911):240–244. https://doi.org/10.1126/science.1164363
2. Tshiala MF, Olwoch JM, Engelbrecht FA. Analysis of temperature trends over Limpopo Province, South Africa. J Geogra Geol. 2011;3(1):13–21. https://doi.org/10.5539/jgg.v3n1p13
3. Attribution of extreme weather events in the context of climate change. Washington DC: National Academic Press; 2016. https://doi.org/10.17226/21852
4. De Cian E, Lanzi E, Roson R. Seasonal temperature variations and energy demand. Clim Change. 2013;116(3–4):805–825. https://doi.org/10.1007/s10584-012-0514-5
5. Fitchett J, Robinson D, Hoogendoorn G. Climate suitability for tourism in South Africa. J Sustain Tour. 2016;25:851–867. https://doi.org/10.1080/09669582.2016.1251933
6. Ziervogel G, New M, Van Garderen EA, Midgley G, Taylor A, Hamann R, et al. WIREs. Clim Change. 2014;5:605–620. https://doi.org/10.1002/wcc.295
7. Lazenby MJ, Landman WA, Garland RM, DeWitt DG. Seasonal temperature prediction skill over southern Africa and human health. Meteorol Appl. 2014;21:963–974. https://doi.org/10.1002/met.1449

8. Cohen JL, Furtado JC, Barlow M, Alexeev VA, Cherry JE. Asymmetric seasonal temperature studies. Geophys Res Lett. 2012;39(4):1–7. https://doi.org/10.1029/2011GL050582

9. Kutta E, Hubbart JA. Reconsidering meteorological seasons in a changing climate. Clim Change. 2016;137(3–4):511–524. https://doi.org/10.1007/s10584-016-1704-3

10. Brunetti M, Maugeri M, Monti F, Nanni T. Temperature and precipitation variability in Italy in the last two centuries from homogenised instrumental time series. Int J Climatol. 2006;26(3):345–381. https://doi.org/10.1002/joc.1251

11. Allen MJ, Sheridan SC. Evaluating changes in season length, onset and end dates across the United States (1948-2012). Int J Climatol. 2016;36:1268–1277. https://doi.org/10.1002/joc.4422

12. Dong W, Jiang Y, Yang S. Response of the starting dates and lengths of seasons in Mainland China to global warming. Clim Change. 2010;99:81–91. https://doi.org/10.1007/s10584-009-9669-0

13. Czernecki B, Miętus M. The thermal seasons variability in Poland, 1951-2010. Theor Appl Climatol. 2017;127:481–493. https://doi.org/10.1007/s00704-015-1647-z

14. Hiltunen L, Haukka J, Ruuhela R, Suominen K, Parton T. Local daily temperatures, thermal seasons, and suicide rates in Finland from 1974 to 2010. Environ Health Prev Med. 2014;19:286–294. https://doi.org/10.1007/s12199-014-0391-9

15. Vigaud N, Pohl B, Crétat J. Tropical-temperate interactions over southern Africa simulated by a regional climate model. Clim Dyn. 2012;39:2895–2916. https://doi.org/10.1007/s00382-012-1314-3

16. Tyson PD, Preston-Whyte RA. The atmosphere and weather of southern Africa. 2nd ed. Cape Town: Oxford University Press Southern Africa; 2000. https://doi.org/10.1002/joc.725

17. Engelbrecht C, Landman W, Engelbrecht F, Malherbe J. A synoptic decomposition of rainfall over the Cape south coast of South Africa. Clim Dyn. 2015;44(9–10):2589–2607. https://doi.org/10.1007/s00382-014-2230-1

18. Roffe SJ, Fitchett JM, Curtis, CJ. Classifying and mapping rainfall seasonality in South Africa: A review. S Afr Geogr J. 2019;101(2):158–174. https://doi.org/10.1080/03736245.2019.1573151

19. Doran JD. Regional climate messages: Southern Africa. Scientific report from CARIAA Adaption at Scale in Semi-Arid Regions (ASSAR) Project. Unpublished report; 2014.

20. Davis-Reddy CL, Vincent K. Climate risks and vulnerability: A handbook for southern Africa. 2nd ed. Pretoria: CSIR; 2017.

21. Dyson LL, Van Heerden J. A model for the identification of tropical weather systems (MITS) over South Africa. Water SA. 2002;289(3):249–258.

22. Lennard C, Hegerl G. Relating changes in synoptic circulation to surface rainfall response using self-organising maps. Clim Dyn. 2014;44(3–4):861–879. https://doi.org/10.1007/s00382-014-2169-6

23. Weather SA. Weather questions [webpage on the Internet]. No date [cited 2019 Aug 20]. Available from: http://www.weathersa.co.za/home/weatherques

24. Alpert P, Osentinsky I, Ziv B, Shafir H. A new seasons definition based on classified daily synoptic systems: An example for the Eastern Mediterranean. Int J Climatol. 2004;24(8):1013–1021. https://doi.org/10.1002/joc.1037

25. Trenbreth KE. What are seasons? Bull Am Meteorol Soc. 1983;64(11):1276–1282. https://doi.org/10.1175/1520-0477(1986)064<1276:WATS>2.0.CO;2

26. Karl TR, Jones PD, Knight R, Kukla G, Plummer N, Razuvzyev V, et al. Asymmetric trends of daily maximum and minimum temperature. Bull Am Meteorol Soc. 1993;74(6):1007–1023. https://doi.org/10.1175/1520-0477(1933)074<1007:anporg>2.0.co;2

27. Klopper E, Landman WA, Van Heerden J. The predictability of seasonal maximum temperature in South Africa. Int J Climatol. 1998;18(7):741–758. https://doi.org/10.10029(sici)1097-0088(19980615)18:7<747::aid-joc279>3.0.co;2-b

28. Kruger AC, Shongwe S. Temperature trends in South Africa: 1960–2003. Int J Climatol. 2004;24(15):1929–1945. https://doi.org/10.1002/joc.1096

29. Kruger AC, Nxumalo M. Surface temperature trends from homogenized time series in South Africa: 1931–2015. Int J Climatol. 2017;27(5):1–52. https://doi.org/10.1002/joc.4851

30. Benhin JK. Climate change and South African agriculture: Impacts and adaption options [document on the Internet]. c2006 [cited year month day]. Available from: https://www.weadapt.org/sites/weadapt.org/files/legacy-new/knowledge-base/files/5370f181a5657504721bd5c21csouth-african-agriculture.pdf

31. Craig MH, Kleinschmidt I, Nawn JB, Le Sueur D, Sharp BL. Exploring 30 years of malaria case data in KwaZulu-Natal South Africa: Part 1. The impact of climatic factors. Trop Med Int Health. 2004;9(12):1247–1257. https://doi.org/10.111/j.1365-3156.2004.01340.x

32. Whitecross MA, Witkowski ETF, Archibald S. Savanna tree-grass interactions: A phenological investigation of green-up in relation to water availability. S Afr J Bot. 2017;108:29–40. https://doi.org/10.1016/j.sajb.2016.09.003

33. Fitchett JM, Grab SW, Portwig H. Progressive delays in the timing of sardine migration in the southwest Indian Ocean. S Afr J Sci. 2019;115(7/8), Art. #5887, 6 pages. https://doi.org/10.17159/sajs.2019/5887

34. Mackellar N, New M, Jack C. Observed and modelled trends in rainfall and temperature for South Africa: 1960–2010. S Afr J Sci. 2014;110(7/8), Art. #2013-0353, 13 pages. https://doi.org/10.1590/sajs.2014/20130353

35. Government Communications (GCIS). South Africa yearbook 2015/2016. 23rd ed. Pretoria: Government Communications; 2016.

36. Dube LT. Climate of southern Africa. S Afr Geogr J. 2002;84(1):125–138. https://doi.org/10.1080/03736245.2002.9713763

37. Nicholson SE. The ITCZ and the seasonal cycle over the equatorial Africa. Bull Am Meteorol Soc. 2018;99(2):337–348. https://doi.org/10.1175/BAMS-D-16-0287.1

38. Kruger AC, Goliger AM, Retief JV, Sekele S. Strong wind climatic zones in South Africa. Wind Struct An Int J. 2010;13(1):37–55.

39. Engelbrecht F, Landman W. South Africa's present-day climate. In: Southern Africa risk and vulnerability atlas. Pretoria: CPD Print; 2010. p. 2–5.

40. Davis CL. Climate risk and vulnerability: A handbook for southern Africa. Pretoria: CSIR; 2011.

41. Unal Y, Kindap T, Karaca M. Redefining the climate zones of Turkey using cluster analysis. Int J Climatol. 2003;23:1045–1055.

42. Bocard D, Gillet F, Legendre P. Numerical ecology with R. New York: Springer; 2011. https://doi.org/10.1007/978-1-4419-7976-6

43. Wilks DS. Cluster analysis. Int Geophys. 2011;100:603–616. https://doi.org/10.1002/joc.1376

44. Kodinariya TM, Mwakwana PR. Review on determining number of clusters in K-means clustering. Int J Adv Res Comput Sci Manag Stud. 2013;1(6):90–95.

45. Studer M. WeightedCluster Library Manual: A practical guide to creating typologies of trajectories in the social sciences with R. LIVES Working Papers. 2013;24:1–32. https://doi.org/10.12682/lives.2296-1658.2013.24

46. Kaufman L. Rousseeuw PJ. Finding groups in data. An introduction to cluster analysis. New York: John Wiley and Sons; 1990. https://doi.org/10.1002/9780470316801

47. Li R, Nathan JBK, Yang J, Wang Y. A phylogenetically informed delineation of floristic regions within a biodiversity hotspot in Yunnan, China. Sci Rep. 2015;5:1–7. https://doi.org/10.10.1038/sreps09396

48. Crow TM, Albeke SE, Buerkle CA, Hufford KM. Provisional methods to guide species specific see transfer in ecological restoration. Ecosphere. 2018;9(1):1–14. https://doi.org/10.1002/ecs2.2059

49. Carvalho PR, Munita CS, Lapolli AL. Validity studies among hierarchical methods of cluster analysis using cophenetic correlation coefficient. Braz J Radiat Sci. 2019;7(2A), Art. #668. https://doi.org/10.15392/bjrs.v7i2a.668

50. Zhou Y, Zhang J. Application of GIS in downscaling regional climate model results over the province of Ontario. Environ Syst Res. 2014;3(1):8. https://doi.org/10.1189/2193-2697-3-8

51. Contractor S, Alexander LV, Donat MG, Herold N. How well do gridded datasets of observerd daily precipitation compare over Australia? Adv Meteorol. 2015, Art. #325718, 15 pages. https://doi.org/10.1155/2015/325718

52. Scott AA, Waugh DW, Zaitchik BF. Reducing urban heat island intensity under warmer conditions. Environ Res Lett. 2018;13(6), 064003. https://doi.org/10.1088/1748-9326/aabd6c

53. National Oceanic and Atmospheric Administration (NOAA). Anomalies vs temperature [webpage on the Internet]. c2020 [cited 2019 Sep 17]. Available from: https://www.ncdc.noaa.gov/monitoring-references/dyk/anomalies-vs-temperature

54. Roffe SJ. A statistical analysis of rainfall seasonality across South Africa [thesis]. Johannesburg: University of the Witwatersrand; 2019.

55. Rana S, McGregor J, Renwick J. Precipitation seasonality over the Indian subcontinent: An evaluation of gauge, reanalyses, and satellite retrievals. J Hydromet. 2015;16:631–651.

56. Tshiala MF, Olwoch JM. Impact of climate variability on tomato production in Limpopo Province, South Africa. Afr J Agric Res. 2010;5(21):2945–2951.

AUTHORS:
Charity R. Nhemachena[1]
Binganidzo Muchara[1]

AFFILIATION:
[1]Department of Finance and Economics, Graduate School of Business Leadership, University of South Africa, Pretoria, South Africa

CORRESPONDENCE TO:
Charity Nhemachena

EMAIL:
nhemachenacharity@gmail.com

DATES:
Received: 17 Feb. 2020
Revised: 23 May 2020
Accepted: 28 May 2020
Published: 29 Sep. 2020

HOW TO CITE:
Nhemachena CR, Muchara B. Structure of the sunflower plant breeders' rights landscape in South Africa. S Afr J Sci. 2020;116(9/10), Art. #7966, 6 pages. https://doi.org/10.17159/sajs.2020/7966

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:
varieties, breeder, research, plant

FUNDING:
None

Structure of the sunflower plant breeders' rights landscape in South Africa

Varietal innovations and protection of plant breeders' rights (PBRs) contribute to the development of any crop's ability to produce higher yields relatively consistently. Producing yields under adverse weather conditions and the overall characteristic of drought tolerance, make the sunflower an attractive crop for producers in dryland production regions. The main objective of this study was to give an overview of the structure of the South African sunflower breeding programme, focusing on the construction of PBRs and the leading players in sunflower breeding and seed production in South Africa. We compiled a detailed database of sunflower varietal innovations in South Africa from 1979 to 2019 using various sources such as the South African Grain Laboratory, the Department of Agriculture's Plant Variety Journals and the Crop Estimation Committee. This data set was then analysed using descriptive statistics and trend analysis to determine the main trends in ownership of PBRs and sunflower varieties. We looked at the inclusion of new sunflower varieties on the national variety list for sunflower varietal improvements in South Africa over this period. A total of 76 PBR sunflower varietal applications were lodged for the period – an average of 1.9 applications per year. The principal applicants for varietal inclusions on the national variety list were Pannar with 102 varieties (23.8%), Pioneer seeds with 51 varieties (11%), Saffola seed with 42 varieties (9.8%) and Agricultural Research Council with 10 varieties (2.3%). In order for breeders to benefit from their investment in research and avoid exploitation of their work, they need to be protected and receive returns on their investments. Innovation can be stimulated by proper collaboration between the private and public sectors, aided by broader variety sector legislation that encourages all players to invest.

Significance:

- The study addresses the absence of empirical proof on the patterns and trends of sunflower varietal improvements in South Africa.
- The study provides evidence on the evolution of varietal rights, the extent of varietal rights granted, and which sunflower varieties were included on the national variety list and the breeders and owners of the varieties.

Introduction

There is increased investment from the private sector in the agriculture and seed industry in South Africa, which has triggered a significant debate on intellectual property rights.[1] Several legislations have been formulated at a national and international level on intellectual property rights. The protection of intellectual property rights of plant breeders was recognised in the 19th century. The International Union for the Protection of New Varieties of Plants (UPOV) was established in 1961 and seeks to coordinate plant variety protection laws and standards of protection across member countries. Plant variety protection allows exclusively the protection granted for plant-related innovations.[1]

South Africa became the tenth member of UPOV in 1977. In 2001, there were more than 50 member countries, and the number increases every year. The main objectives of UPOV are to bring about standard laws on plant breeders' rights, to standardise procedures for the testing of new varieties and to promote cooperation between member countries. The advantage of being a UPOV member is the privilege of any person within a member country to apply for plant breeders' rights in any other member country. Plant breeders' rights (PBRs) are a form of intellectual property rights that are valid only in the country in which they were granted.[2] PBRs can be granted to a variety in different countries. Since 1996, the country has not acceded to UPOV 1991 through amendment of the *Plant Breeders' Rights Act*[3] which was enacted in 1976[4]. PBR provides for the acquisition of legal rights in terms of this *Plant Breeders' Rights Act*, to obtain royalties as a return on research efforts and investments made during the process of breeding of a new plant variety. The whole process provides the owner of a variety the opportunity to obtain a financial reward for their efforts, as the breeding and development of a new variety are expensive and time consuming. It is important to develop new and improved plant varieties as there is a continuous demand for better quality, higher yields, better processing properties and increased disease resistance.

A PBR is valid for a term of 20 or 25 years, depending on the type of plant. During the first 5 to 8 years (the period of sole right for the breeder), the owner has the sole right to multiply and market propagating material of the variety. During the next 15 years, the holder is required to issue licences to other persons who also desire to use the material. If the holder of the right refuses to issue licences, these individuals may apply to the Registrar for a compulsory licence.[2] During the term of the right, the holder may continue to claim royalties from all licensees for any propagating material produced and sold. It is only after the expiration of the full term of the plant breeder's right, that the variety becomes openly accessible to the public, and anyone may then propagate and sell it.[2] Foreign breeders and variety owners are not keen to supply propagating material to individuals in other countries if such material cannot be protected under PBR. Hitherto, very little foreign propagating and breeding material would be available in South Africa if we did not have a PBR system.[2] Pardey et al.[5] argue that plant varietal rights are subject to ongoing public policy scrutiny and debate. To inform these policy discussions, there is a need to understand the

evolution of varietal rights and the extent of the varietal rights granted. It is also essential to follow changes in the rights on offer over time such as changing ownership of the rights (including a comparison between public and private as well as domestic and foreign breeders) and the impact of plant variety protection on varietal development.

In this study, we focused on sector analysis in the South African sunflower industry regarding intellectual property rights in the form of PBR and the leading role players in sunflower breeding and production in South Africa. We provide insight into the pattern of application for and granting of PBR and also the inclusion of a variety on the national variety list, with a focus on the main players in the sector. The trend analysis of PBR application for the period 1979–2019 helps to define the technology trend in the application of PBR and the registration sunflower varieties on the national variety list. The analysis further determines the focus of local public and private, and foreign public and private sector research through the inclusion of sunflower varieties on the national variety list.

An overview of the sunflower industry

Sunflower is a native wild crop from North America which originated around the Fertile Crescents or South and Central America, and native Americans used it as a food source. In North America, it was crushed or pounded into flour for cakes and bread.[6,7] Europeans were exposed to sunflower cultivation from places like southern Canada and Mexico to Spain. Spain was the first gateway of sunflowers into Europe, and then the crop spread until it reached Russia.[7] It was in Russia where the first commercial production started.[6] In the early 19th century, Russian farmers were growing over 810 000 hectares of sunflower seed. Russia began to produce high oils from sunflower seed in 1860 and this increased oil content from 28% to almost 50%. These oils from Russia were exported into the USA after World War II, which stirred world interest in the crop.[7] The first commercial hybrid of sunflower was released in 1972, and it produced a 25% higher yield than the normal variety.[8] The genetic modification allowed for the release of short-stemmed and high-yield varieties that enabled efficient mechanised cropping, resulting in sunflower being a major oilseed crop.[8]

In South Africa, sunflower is listed as the third most important crop after maize and wheat, and the Crop Estimation Committee reported a yield of 820 000 tons in the 2017/2018 production season.[9] Sunflower has been classified as an ideal crop because of its characteristics of thriving under marginalised and low-input farming conditions. Sunflower has the ability to produce consistent yields under hostile weather conditions. It is drought tolerant and hence attractive to dryland farmers. The crop can grow in marginal soils with little or no fertiliser. In South Africa, it is regarded as a low input and beneficial cash crop following maize.[10] Sunflower can take advantage of residual nitrogen and subsoil moisture that is underutilised by other crops; hence, it is a useful crop for crop rotation.

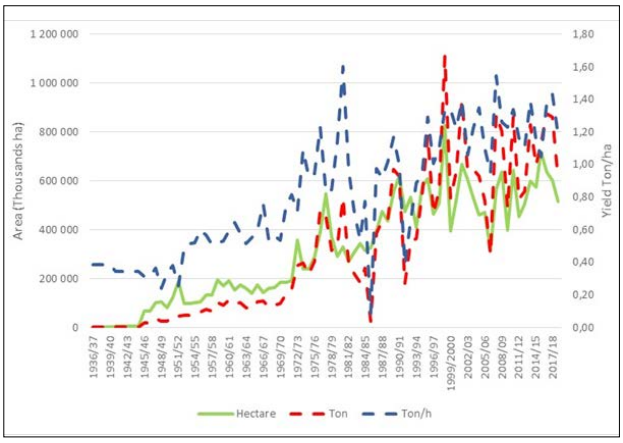
Sunflower seed is mainly used for human consumption for oil production and cake feed for animals, especially in the dairy industry.[9] Although the whole seed is utilised, sunflower constitutes about 10–20% of the volumes of crops produced in South Africa each year. The prices for confectionery sunflower seed are significantly higher than sunflower seed that is used for oil processing. Some confectionery sunflower is processed into foods such as granola bars, healthy multigrain breads, roasted as a snack or for other baking purposes. The sunflower seed also produces oil cake, that is mainly used for animal feeds as sunflower oilcake meal because of its high protein content. Sunflowers can be used as an alternative for peanut butter to make sun butter for people with nut allergies. It is also mixed with rye flour to make bread (in Germany, this is called Sonnenblumenkernbrot – literally: sunflower whole seed bread).[6] Sunflower seeds are sold as food for birds and can be used directly in cooking and salads. Sunflower and soybeans are important sources of oils and protein for both household and industrial purposes.[11] When seeds are dehulled, the seed can be consumed or made into different oils. Sunflower oil is used daily by restaurants, food industries and households as cooking oil, margarine and spreads. Sunflower oil can also be converted into biodiesel.[12]

Sunflower also has non-consumption uses, such as purple dye for textiles, body painting and cosmetics. The other various parts of the plant are used medicinally for the skin and hair, and the dried stalk is used as building material. The plant is also used as an ornamental flower in Europe.[6] Sunflowers are also used to produce latex; they are a subject of experiments to improve their suitability as an alternative crop for producing hypoallergenic rubber.

Production areas of sunflower in South Africa

Sunflower seed is produced in six of the nine provinces in South Africa. The Free State and North West Provinces are the major producers of sunflower seed, followed by Limpopo, Mpumalanga and Gauteng.[2] Of the area planted with sunflower in the 2011–2016 period, 54% was in the Free State, 34% in the North West, 11% in Limpopo and 1% in Mpumalanga.[13] The area planted with sunflower increased in the drought-affected season of 2016 due to sunflower's better performance in dry conditions and its late planting window relative to maize.[10]

Figure 1 shows trends in sunflower area and yield in South Africa between 1936 and 2018. The patterns indicate an overall increasing trend in both area and production. The hectares planted for sunflower seed have been volatile for the past 10 years, with an average annual growth of only 1.8%. Only in recent seasons have the number of hectares planted stabilised at around 600 000 hectares per annum. The sunflower crop yield of the 2017/2018 season, as shown in Figure 1 is over 600 000 tons. The crop decreased by 1.4% (12 000 tons) in this period. The major sunflower-producing provinces, namely the Free State and North West, contributed 95% of the total crop.



Source: Data from SAGIS[17]

Figure 1: Commercial sunflower area, production and productivity in South Africa, 1936–2019.

Research methods and data

Plant breeders' rights with respect to sunflowers in South Africa were analysed to assess the sources of intellectual property rights. We used secondary data collected from various sources on sunflower variety policies and changes in sunflower PBRs. Some studies analysed changes in plant variety protection focusing on trends and changes in plant variety including protection policies.[5,14,15] These studies analysed trends to understand the evolution of plant varietal protection. We applied the same approach in order to understand changes that have shaped the South African sunflower varietal improvement landscape to date. The findings allow comparison with other countries in future.

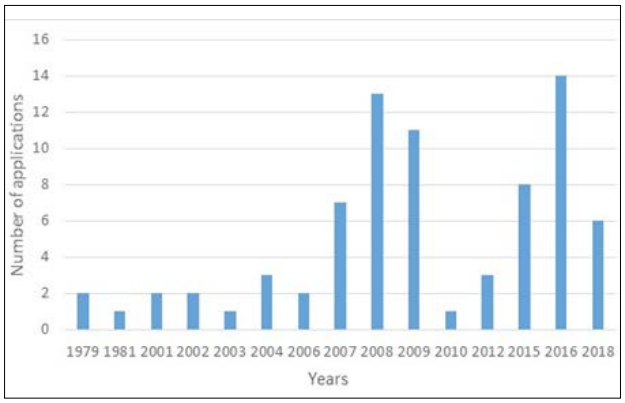
Data on sunflower varietal innovations in South Africa from 1979 to 2019 were collected from the Plant Variety Journal which is compiled under the supervision of the Registrar of Agriculture in the South African Department of Agriculture, Forestry and Fisheries. The database provides information to assess changing amounts and forms of sunflower PBRs as well as changes in the types of applicants of the rights and inclusion of the South African seed variety list. The database includes data on applications and granting of the PBRs and applications for additions on

the national variety list. Further information captured in the database includes plant variety name, alias name, applicant name, applicant type, application date of PBR, grant date of PBR and application date for the varieties to be included on the national variety list. A supplementary file was provided to show the principal applicants and the sunflower varieties they own.

Overview of sunflower varietal releases in South Africa

Trends of sunflower inclusion on the national varietal list

Figure 2 shows the trends of annual applications for PBRs for sunflower varieties lodged in South Africa. The total number of PBRs submitted for sunflower varieties were 76 during the period from 1979 to 2019 – an average of 1.9 applications per year. The period starts in 1979 because that is the year that PBR applications began to be recorded in the Plant Variety Journal. Data on the number of sunflower PBRs lodged since 1979 show changes over time, with some years having a high number of applications while others recorded very low applications. The highest number of applications (14 in total) was made in 2016. There were no applications from 1979 up to 1980 and between 1982 and 2000. Furthermore, no applications were lodged in 2005, 2011, 2013, 2014 and 2017. It can be argued that the gaps could be because most breeders and seed sellers focused on registering their sunflower varieties on the national variety list. Applications for sunflower varietal inclusions on the national variety list are typically recorded in the Plant Variety Journal before application for new varieties. The total number of PBRs lodged for sunflower varieties were compared to maize and wheat which have a very high number of PBRs applications.

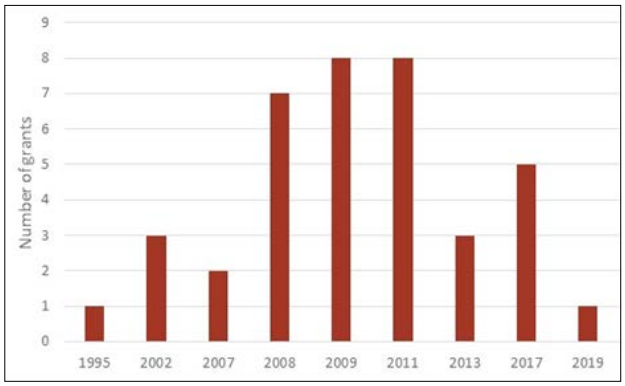


Source: Data obtained from Plant Variety Journal

Figure 2: Annual applications for sunflower varietal improvements by plant breeders, 1979–2019.

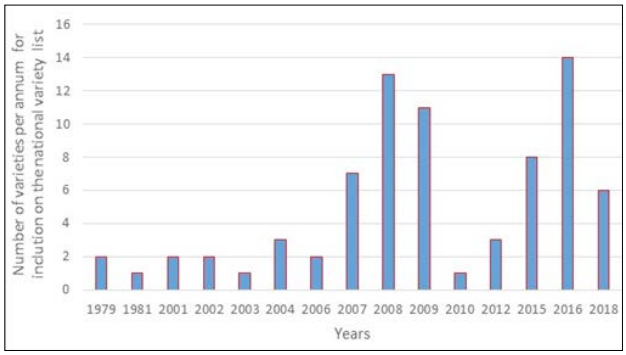
Figure 3 presents an analysis of the PBRs for sunflower varieties that were granted. A total of only 38 PBRs grants were approved in the period 1979–2019, with the first approval in 1995. The political situation before 1994 could have contributed to this low number as most of the applicants were private companies partnering with foreign companies. The highest number of PBRs were granted in 2009 and 2011 with both years recording nine grants. The average waiting period between application for and granting of protection is more than 365 days and this overall period may be threatening innovation in this sector.

Figure 4 shows the applications that were lodged by breeders and seed companies for their sunflower varieties to be included on the national seed varieties list. A total of 429 applications were submitted for inclusion of sunflower varieties on the national variety list in the study period. The highest number of applications was recorded in 1991 with approximately 38 applications and the lowest number of applications (2) was recorded in 2010. In the sunflower industry, there are more applications from plant breeders for inclusion on the national variety list than for protection of PBRs.



Source: Data obtained from Plant Variety Journal

Figure 3: Plant breeders' rights granted for sunflower varieties, 1979–2019.



Source: Data obtained from Plant Variety Journal

Figure 4: Annual applications for the inclusion of sunflower varieties on the national variety list, 1979–2019.

It is important to note that, despite no new applications being lodged between 1982 and 2000 (Figure 2), there were large numbers of applications for inclusion of varieties in the national list for the same period (Figure 4). Breeders were therefore focusing on listing their varieties instead of continuing to produce new varieties whose listing on the national varietal list was not certain. The registration of new varieties on the national list is driven by government entities, while breeding is mostly done by private companies. Synchronisation of the activities and processes is important to minimise delays in the processing of applications as well as granting of PBRs. Often breeders are directly affected by the delays in the registration system with potential revenue losses. The knock-on effect is that breeders are discouraged from continued investment in new varieties and consequently the sector, as well as the economy, is negatively affected.

Applicants for inclusion of sunflower varieties on national variety list

We analysed the nature of applicants seeking inclusion of sunflower varieties on the South Africa national variety list for the period 1979–2019. The focus was on identifying the changing public, private and foreign owners' roles in the sunflower national variety list. Table 1 presents the composition of applicants (breeders and seed sellers) of sunflower varieties on the national variety list. Based on analysis of shares of sunflower varieties included on the national variety list since the first publication of the South African Plant Variety Journal in 1979, there were 54 applicants and the main applicants were Pannar with 102 varieties (23.8%), Pioneer seeds with 51 varieties (11%), Saffola seed with 42 varieties (9.8%) and the Agricultural Research Council (ARC) with 10 varieties (2.3%). The public sector represented by ARC came into play to apply for inclusion of their sunflower varieties on the national variety list in 1995, which shows that the public sector was not dealing much with sunflower. The main applicants and also owners of sunflower varieties are private companies in South Africa and foreign private companies that collaborate with local private companies. These foreign

companies are Syngenta South Africa and Syngenta France, Cargill SA and Cargill USA and Pioneer Seeds and Sipco Sun Products. In a study done in the Netherlands, Louwaars et al.[16] found that the public sector (universities, government bodies and private non-profit organisations) submitted 23.8% of plant based patents, while in a study done in the USA, public bodies were granted 21.9%. Furthermore, Louwaars et al.[16] found that in the Netherlands, private companies dominated the number of plant-based patents granted in the country. In South Africa, the wheat sector is also governed by the private sector with companies like Sensako and Pannar.[15]

Table 1: Applicants for inclusion on the national variety list

Applicant	Number of applications	Percentage of applications
Cereal Research Non-Profit Company and Sanniesguns Boerdery Trust	2	0.5
Abeseed	4	0.9
Abeseed and Interstate Seed Co USA	2	0.5
Advanta Seeds England	3	0.7
Advanta Seeds SA and Argentina	4	0.9
Advanta Seeds SA and Interstate Seed USA	2	0.5
Africa Pacific Seeds	2	0.5
Agricol	7	1.6
Agricol and Lefroy Seeds Australia	5	1.2
Agricol and Nuseed	4	0.9
Agricol and Seed 2000	1	0.2
Agricol and Triumph Seed USA	4	0.9
Agseed	2	0.5
Agricultural Research Council	10	2.3
Asbseed	4	0.9
Asgrow SA (Pty) Ltd	20	4.7
Bioseeds SA and Bioseeds USA	3	0.7
Cargill SA (Pty) Ltd	8	1.9
Cargill SA (Pty) Ltd and Cargill Argentina	2	0.5
Cargill SA (Pty) Ltd and P.A.G Seeds USA	1	0.2
Carnia	30	7.0
Ciba-Glegy	1	0.2
Continental South Africa and Pacific Seeds	1	0.2
Dr Knapp	1	0.2

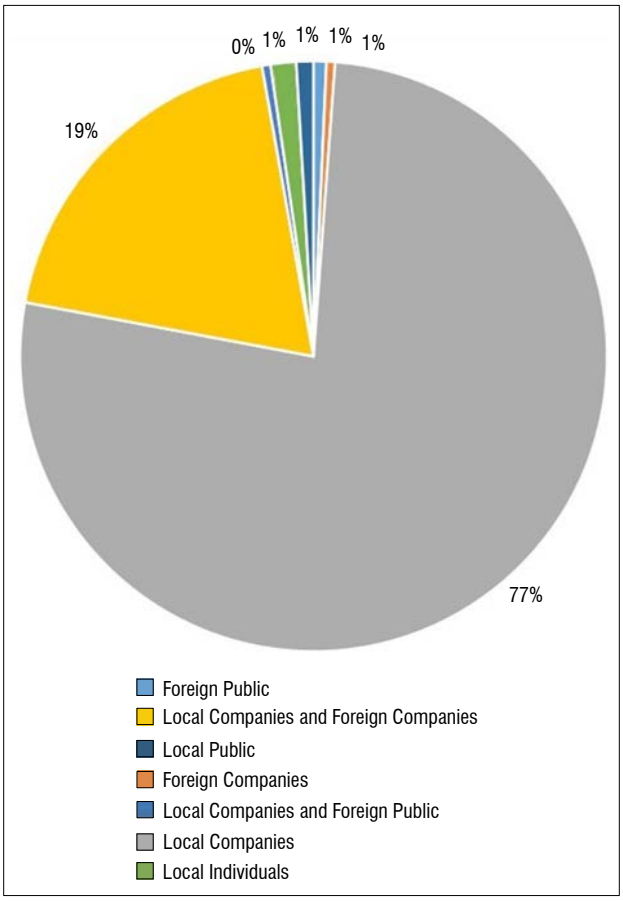
Table 1 continues in the next column

Table 1 continued

Applicant	Number of applications	Percentage of applications
Gunson S.A (Pty) Ltd and Northup King	1	0.2
Gurson SA (Pty) Ltd	2	0.5
H.M Muller	2	0.5
Jackson M., Maurona Flora	3	0.7
Monsanto SA and Monsanto Argentina	4	0.9
Nidera SA and Kksm	4	0.9
P.A.G Seeds	3	0.7
P.A.G Seeds SA and P.A.G Seeds USA	1	0.2
Pannar	102	23.8
Pannar and Dow Gro Sciences Argentina	1	0.2
Pannar and Seed 2000	3	0.7
PHI	5	1.2
Pioneer Hi-Bred, USA	2	0.5
Pioneer Seed	51	11.9
Pioneer Seeds and Sigco Sun Products	28	6.5
RN France and Agricol	1	0.2
Sabi (Edms) Bpk	5	1.2
Saffola Seed	42	9.8
Saffola Seed and Seedtec International USA	4	0.9
Safora Seed (Pty) Ltd	4	0.9
Seedcor	3	0.7
Sensako	21	4.9
Sensako and De Kalb-Phizer	1	0.2
Stauffer Seeds	3	0.7
Sygenta SA and Sygenta France	7	1.6
Thuleka	1	0.2
United Seeds SA and Gabonater Meszte'si Kutató Kht Hungary	1	0.2
Wickens Beleggings BK	1	0.2

Source: Data obtained from Plant Variety Journal

Figure 5 presents the changing composition of applicants of sunflower varietal inclusion on the national variety list. The majority of the varietal inclusion applications for sunflower varieties were filed by local private companies like Pannar, Pioneer, Sensako, Cargill, Carnia, Agroseed and more.



Source: Data obtained from Plant Variety Journal

Figure 5: Applications by local companies, public sector and foreign companies, 1979–2019.

The local private sector constituted the largest share, accounting for 77% of varietal inclusion applications on the national variety list. We further found that the percentage of collaborations between local companies and foreign companies as the second-highest applications, at 19%. Partnerships are formed with local companies when they apply for foreign sunflower varieties to be included on the national variety list on behalf of a foreign company. These companies include Syngenta USA, Syngenta France, Pioneer Hi Brid USA, Dow Grow Sciences Argentina and Bio seed USA. In the South African sunflower industry, the public sector, especially the ARC, continues to play a minor role in the development of sunflower varieties with only a 1% share in the applications of varietal inclusions on the national variety list, and an indication of less activity from the public sector in sunflower production.

The future of the sunflower industry in South Africa

The ARC is exploring the use of crop breeding and engineering principles to produce advanced biofuel. There are various crops identified in South Africa for biodiesel and renewable diesel that include sunflower, soybean and canola. The ARC has been exploiting selected cultivars of sunflower for production of renewable (green) diesel, which is known to meet the requirements for use in a diesel engine. South Africa's White Paper on Renewable Energy highlights the aim of achieving 2–5% levels of biofuels in the national liquid fuel supplies, which currently stand at 400 million litres per annum.[12] Sunflower, canola and soybeans have since been proposed for the production of biodiesel in the country. Biodiesel from plant oils is widely recommended to be an ecofriendly, renewable energy source which can be a suitable replacement for petroleum fuel-derived diesel. Sunflower is one of the leading oil crops in South Africa and one of the biofuel mandates by the government for biodiesel production. Indeed, mandatory blending with renewables might become applicable at various filling stations in South Africa.[12] However, the success of sunflower-dependent industries depends on the success and level of innovation in the release and registration of new sunflower varieties in South Africa. Hybrid varieties are often more productive to meet the ever-growing demand in the renewable energy sector, hence a more efficient seed breeding programme, variety registration and granting of breeders' rights is required to ensure long-term growth and sustainability of the sector.

Conclusion and recommendations

We analysed the evolving landscape of sunflower plant breeders' rights to address the absence of empirical proof of the patterns and trends of sunflower varietal improvements in South Africa. The aim of the study was to provide evidence on the evolution of varietal rights, the extent of varietal rights granted, and which sunflower varieties were included on the national variety list and the breeders and owners of the varieties. The changes of the rights over time, changing ownership of PBRs (including comparison between public and private as well as domestic and foreign breeders) and the impact of plant variety protection on sunflower development were analysed. We put together a detailed and novel count and attribute database of sunflower varietal innovations in South Africa from 1979 to 2019, using information from Plant Variety Journal, Department of Agriculture Forestry and Fisheries, South African National Library and ARC. The empirical analyses were based on descriptive statistics, trend analysis and graphical representation of trends and ownership of PBRs of sunflower varietal improvements and applications of sunflower for inclusion of varieties on the national variety list.

A total of 76 PBR sunflower varietal applications were lodged during the period under study with an average of 1.9 applications per annum. The breeders of new varieties need to be protected for them to receive returns on their investments into varietal improvement research. This can be done by improving the efficiency of the registration and processing of applications related to PBRs. Furthermore, innovations can be stimulated by strong collaborations between the private and public sectors, which can be achieved by improving legislation that governs all players in the sector.

Acknowledgements

We thank the anonymous reviewers who provided comments on this manuscript.

Competing interests

We declare that there are no competing interests.

Authors' contributions

C.R.N. and B.M. were jointly responsible for the paper.

References

1. Srivastava M, Chaudhary V, Pilania D. Intellectual property rights on plant varieties in India: A sector-wise analysis [document on the Internet]. c2015 [cited 2019 Jun 19]. Available from: http://nopr.niscair.res.in/handle/123456789/31288
2. South African Department of Agriculture Forestry and Fisheries (DAFF). Plant Breeders' Rights policy [webpage on the Internet]. c2011 [cited 2019 Jun 27]. Available from: https://www.gov.za/services/plant-production/plant-breeders-rights?
3. Nhemachena C, Kirsten J, Muchara B. The effects of Plant Breeders' Rights on wheat productivity and variety improvement in South Africa. Paper presented at: Agricultural Economics Association of South Africa (AEASA) Annual Conference; 2018 September 25–27; Cape Town, South Africa. https://doi.org/10.22004/ag.econ.284783
4. Moephuli S, Moselakgomo M, Phehane V. The Agricultural Research Council's role in plant variety protection and technology transfer - promoting publicly funded research. Paper presented at: UPOV Workshop; 2012 March 05; Zanzibar, Tanzania.
5. Pardey P, Koo B, Drew J, Horwich J, Nottenburg C. The evolving landscape of plant varietal rights in the United States, 1930–2008. Nat Biotechnol. 2013;31(1):25–29.
6. Boshoff ID. The factors that influence the price of sunflower in South Africa [thesis]. Potchefstroom: North-West University; 2008.

7. South African Department of Agriculture, Forestry and Fisheries (DAFF). Sunflower production guideline 2010 [document on the Internet]. c2010 [cited 2019 May 19]. Available from: https://www.nda.agric.za/docs/Brochures/prodGuideSunflower.pdf

8. Blamey FPC, Zollinger RK, Schneiter AA. Sunflower production and culture. In: Schneiter AA, editor. Sunflower technology and production. Volume 35. Madison, WI: The American Society of Agronomy; 2015. p. 1–19. https://doi.org/10.2134/agronmonogr35.c12

9. Ma'ali S. Delayed planting impacts sunflower yield [webpage on the Internet]. c2017 [cited 2019 June 19]. Available from: https://www.grainsa.co.za/delayed-planting-impacts-sunflower-yield

10. South African Bureau for Food and Agricultural Policy (BFAP). The South African sunflower complex: A report by the Bureau for Food and Agricultural Policy (BFAP). Compiled for the Oilseed and Protein Advisory Development Trust and Oilseed Advisory Committee (OAC). Pretoria: BFAP; 2015 [cited 2019 June 16]. Available from: https://www.bfap.co.za/the-south-african-sunflower-complex/

11. Dlamini TS, Magingxa L, Liebenberg F. Estimating the economic value of the national cultivar trials in South Africa: A case for sorghum, sunflower, soybeans and dry beans. Paper presented at: The Fourth Congress of the Italian Association of Agricultural and Applied Economics (AIEAA); 2015 June 11–12; Ancona, Italy. https://doi.org/10.22004/ag.econ.207288

12. South African Agricultural Research Council (ARC). Annual report. Pretoria: ARC; 2018. Available from: https://www.arc.agric.za/Documents/Annual%20Reports/ARCAnnualReport-OCTOBER2019-low%20res.pdf

13. United States Department of Agriculture (USDA). Southern Africa: Crop production map [webpage on the Internet]. c2020 [cited 2020 May 02]. Available from: https://ipad.fas.usda.gov/rssiws/al/crop_production_maps/safrica/SouthAfrica_Sunflower.gif

14. Srinivasan CS. Concentration in ownership of plant variety rights: Some implications for developing countries. Food Policy. 2003;28(5–6):519–546. https://doi.org/10.1016/j.foodpol.2003.10.003

15. Nhemachena CR, Liebenberg FG, Kirsten J. The evolving landscape of plant breeders' rights regarding wheat varieties in South Africa. S Afr J Sci. 2016;112(3–4), Art. #2015-0164, 8 pages. http://dx.doi.org/10.17159/sajs.2016/20150164

16. Louwaars N, Dons H, Van Overwalle G, Raven H, Arundel A, Eaton DJ, et al. Breeding business. The future of plant breeding in the light of developments in patent rights and plant breeder's rights. CGN Report 2009-14. Wageningen: Centre for Genetic Resources, Wageningen University and Research Centre; 2009. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1720088

17. South African Grain Information Services (SAGIS) [homepage on the Internet]. No date [cited 2019 Jun 23]. Available from: http://www.sagis.org.za/

AUTHORS:
Nothando Dunjana[1]
Rebecca Zengeni[1]
Charity Pisa[1,2]
Menas Wuta[3]
Pardon Muchaonyerwa[1]

AFFILIATIONS:
[1]School of Agricultural, Earth and Environmental Sciences, University of KwaZulu-Natal, Pietermaritzburg, South Africa
[2]Department of Natural Resources Management, Marondera University of Agricultural Sciences and Technology, Marondera, Zimbabwe
[3]Department of Soil Science and Agricultural Engineering, University of Zimbabwe, Harare, Zimbabwe

CORRESPONDENCE TO:
Nothando Dunjana

EMAIL:
ntandodunjana@yahoo.com

DATES:
Received: 22 Jan. 2020
Revised: 10 June 2020
Accepted: 22 June 2020
Published: 29 Sep. 2020

HOW TO CITE:
Dunjana N, Zengeni R, Pisa C, Wuta M, Muchaonyerwa P. Nutrient uptake, yield and taste of oilseed rape (*Brassica napus* L.) and soil chemical properties following amendment with uncomposted and composted tobacco waste and cattle manure. S Afr J Sci. 2020;116(9/10), Art. #7846, 9 pages. https://doi.org/10.17159/sajs.2020/7846

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:
compost, organic amendment, organoleptic test, soil organic carbon

FUNDING:
Organisation for Women in Science for the Developing World; International Foundation for Science (grant C/5764-1)

Nutrient uptake, yield and taste of oilseed rape (*Brassica napus* L.) and soil chemical properties following amendment with uncomposted and composted tobacco waste and cattle manure

The inadequacy of the nutrient supply of most tropical and sub-tropical soils may be curbed through organic material recycling, thus reducing the need for mineral fertiliser use. To promote tobacco waste recycling in a smallholder food-cash crop production system, nutrient uptake, dry biomass yield and taste of oilseed rape (*Brassica napus* L.) and soil chemical properties were determined on a sandy loam soil under field conditions. The experiment was a randomised complete block design with three blocks and eight treatments, namely, control (no amendment), mineral fertiliser (121, 30.8, 24.6 kg/ha N, P and K, respectively), uncomposted tobacco leaf scrap (TSC) and compost of TSC and cattle manure (TSC-CM) at 5, 20 and 40 t/ha. N, P and K uptake and dry biomass yield of oilseed rape were higher ($p<0.05$) than control with 40 t/ha TSC-CM and mineral fertiliser application at 3 weeks after transplanting (WAT), while significant improvements with TSC were observed from 5 WAT. Mineral N, extractable P and exchangeable K were higher than control with TSC-CM at 20 t/ha and 40 t/ha at 3 WAT, and higher with TSC at 9 WAT. Soil organic carbon was more improved with TSC application than TSC-CM at 9 WAT. Organoleptic testing revealed an intensely bitter taste in oilseed rape with mineral fertiliser, 20 t/ha and 40 t/ha TSC at 3 WAT, although it diminished with time. A trade-off of nutrient uptake, yield, taste of oilseed rape and soil properties improvement is attainable with application of TSC-CM at 40 t/ha, while if composting is not feasible, TSC application at 20 t/ha is a viable alternative. Thus, judicious utilisation of tobacco waste offers a viable solution to the problem of low soil fertility on sandy soils and can reduce the need for mineral fertiliser use, while promising sustainable soil management.

Significance:

- The study underlines the importance of recycling of organic materials as a viable, environmentally safe and low-cost soil fertility management practice.
- The study presents pragmatic practices that may be adopted so as to optimise the benefits of use of tobacco waste on food-cash crop smallholder farms, given the increasing popularity of such farming systems in Africa.

Introduction

The cultivation of green leafy vegetables in smallholder farms across sub-Saharan Africa is widespread[1,2] and essential for crop diversification. Crop diversification increases farm productivity, income and food variety while cushioning smallholder farmers against food price and climate shocks.[3] The leaf biomass of *Brassica* provides fibre, vitamins, macronutrients, micronutrients and antioxidants to rural populations, which supplements nutrients obtained from the predominantly cereal-based staple diets.[4] In the southern African context, like in Zimbabwe, oilseed rape (*Brassica napus* L.) is grown for its leaves and is one of the primary vegetable crops due to its high profitability, all-year round production and its short lifespan as compared to other leafy vegetables.[5] However, oilseed rape has high nutrient requirements which can be met through mineral and/or organic fertiliser applications to attain optimum yield.[6] Resultantly, continuous cropping without sufficient nutrient replenishment reduces productivity and poses a serious threat to the sustainable use of soil.

Soil infertility characterised by nutrient deficiencies, soil acidity and declining soil organic matter has been recognised as the fundamental biophysical constraint to crop productivity among smallholder farms,[7] thus various efforts have been directed towards addressing this problem. Ganya et al.[5] showed that nitrogen fertilisation at increasing rates significantly increases oilseed rape biomass yield, while leguminous trees' biomass yields have been tested in isolation and in combination with N-fertiliser with positive results.[8,9] However, the prohibitive prices of mineral fertilisers[10] and risk of soil acidification[11] may discourage its recommendation. Additionally, Maereka et al.[12] found a correlation between nitrate accumulation in *Brassica juncea* leaves and N-fertilisation, which resulted in a bitter taste of the vegetables, which raises the question of the suitability of intensive mineral fertiliser application on quality of vegetables. On the other hand, the adoptability of promising organic amendments including animal manures and agroforestry technologies is constrained by factors including unavailability in adequate quantities as well as variable quality.[8,13,14] Accordingly, site-specific soil fertility management interventions are key in fostering relevance and adoptability.

Soil amelioration with tobacco waste (*Nicotiana tabacum* L.) has been shown to be beneficial in improving the productivity of various horticultural crops including lettuce (*Lactuca sativa*), spinach (*Spinacia oleracea)* and tomato (*Lycopersicon esculentum*).[15-17] Further, soil biological and physical properties have also been enhanced by tobacco waste application in other parts of the world.[18,19] Tobacco waste nutrient levels in the range 1.97–2.38% N, 0.21–0.50% P, 0.32–1.03% K and 38–41% C have been documented,[19,20] pointing to its potential value as a nutrient

resource. Meanwhile, tobacco production in smallholder farming systems has been on the rise in sub-Saharan Africa, particularly in leading tobacco-producing countries, such as Tanzania, Zimbabwe and Malawi, due to the policy reforms in the tobacco value chain and land resettlement initiatives.[21,22] For example, in Zimbabwe, on average, production of tobacco occurs on 1.3-ha plots on the resettlement smallholder farms.[23] Although tobacco production has been on the decrease in developed parts of the world, the economic benefits are prioritised over the health risks in developing countries.[24] Subsequently, tobacco production creates a niche for organic waste generation in the form of tobacco leaf scrap (TSC). Tobacco leaf scrap is generated on-farm during harvesting, grading and packaging of the tobacco leaf. As such, TSC can be beneficially utilised to augment the limited organic resources in smallholder farming systems for the production of food crops, including oilseed rape.

Composting of tobacco waste with different organic materials such as cattle, pig and poultry manure has been widely employed as an environmentally friendly method of tobacco waste handling in tobacco processing industries, while generating a nutrient rich soil ameliorant.[18,25] Composting ensures detoxification of nicotine before soil amelioration[18,26,27] as a high concentration of nicotine is linked to depressed microbial activity[27]. However, labour constraints in smallholder food-tobacco producing farming areas[28] as well as limited access to animal manures may deter adoption of composting technologies. On the other hand, other studies have attested to the positive benefits of direct application of tobacco waste such as favourable electrical conductivity (EC) of soil solution, increased microbial activity due to high carbon substrate, high macronutrient supply, insecticidal properties and improved soil aggregate stability due to phenolic properties.[29-31] Further, Nota et al.[32] and Seckar et al.[33] posit that the risk of nicotine toxicity with raw tobacco application is low due to rapid degradation in soil. As such, there is scope for evaluation of the potential of use of tobacco waste as a soil ameliorant so as to develop recommendations that are relevant across diverse smallholder farms.

Additionally, despite the extensive literature on tobacco waste use in vegetable production, no organoleptic (taste) studies relating tobacco waste application to vegetable taste have been done, yet the study by Maereka et al.[12] attests to the key role that fertilisation plays on vegetable quality. Accordingly, given the limited access to fertiliser inputs and paucity of information on the use of tobacco waste under smallholder systems, we sought to investigate the comparative nutrient uptake, yield and taste of oilseed rape and selected chemical parameters of soil after amendment with uncomposted TSC and compost of TSC and cattle manure (TSC-CM) at increasing rates relative to recommended mineral fertiliser application rate. We hypothesised that TSC and TSC-CM would increase nutrient uptake and yield of oilseed rape and fertility status of the soil relative to the control and mineral fertiliser treatments. We further hypothesised that TSC would increase nitrogen and possibly nicotine uptake causing bitterness in oilseed rape relative to control, mineral fertiliser and TSC-CM treatments.

Materials and methods

Site description

The field experiment was conducted at the Marondera University of Agricultural Sciences and Technology (MUAST) farm, Mashonaland East Province in Zimbabwe. The MUAST farm lies at 18°21'56" S and 31°42'19" E in Svosve farming area. The area is typically characterised by hot, wet summers and cold, dry winters. An annual rainfall of 700–1000 mm is received between the months October to April, marking the wet season, while winter season is experienced between May and August. The dominant soil type is the granite-derived sandy soils (Arenosol),[34] which is typically highly leached and infertile.

Chemical composition of the organic amendments used in the experiment

Tobacco leaf scrap used for the experiment was collected from MUAST tobacco curing barns, grading and packaging sheds, while cattle manure (CM) was carefully swept off the top layers of the MUAST farm cattle pen. The TSC-CM was composted by mixing TSC and CM at a 50:50 w/w basis in 1-m^3 perforated wooden bins. Aeration of compost was done weekly by turning of compost and moisture content was maintained at 60%. After 69 days, compost was deemed mature, with compost temperature matching ambient temperature. The chemical composition of TSC, CM and TSC-CM is given in Table 1.

Field experimentation

Nursery of oilseed rape (var. Hobson) was prepared in October 2017 by drilling seeds into a fine tilth seedbed, and lightly covering with soil, grass mulch and watering. Watering was done using a watering can fitted with a rose sprinkler every 2 or 3 days to ensure adequate moisture. Mulch was removed a week after seedling emergence. Five days prior to transplanting, seedlings were treated with Malathion 50% E.C and Carbaryl 50 WP to control aphids and cutworms. Seedlings were transplanted 28 days after emergence, prior to which they were hardened through water stressing for 7 days.

Seedlings were transplanted onto a sandy loam soil with pH (H_2O) of 5.4, total N, extractable P and soil organic carbon (SOC) content of 0.7 g/kg, 18.7 mg/kg and 5 g/kg, respectively. The field was ploughed to a depth of 0.25 m using a tractor-drawn disc plough followed by levelling using a hand-held hoe. The experiment was laid out in a randomised complete block design with three blocks each consisting of eight plots measuring 3.5 x 3 m. Seedlings were transplanted at a plant density of 83 333 plants/ha. The treatments consisted of a control with no fertility amelioration, mineral fertiliser, TSC and TSC-CM applied at 5, 20 and 40 t/ha. The mineral fertiliser treatment was 500 kg/ha compound D + 250 kg/ha ammonium nitrate supplying equivalents of 121 kg N/ha, 30.8 kg P/ha and 24.6 kg K/ha representing the high nutrition required for oilseed rape as recommended in the growers manual.[6] The organic amendments were spread and incorporated into soil by digging with hand-held hoes to about 0.1 m, a day before transplanting, while basal mineral fertiliser was applied at transplanting. All plots were watered immediately after transplanting. Top dressing fertiliser was applied to the mineral fertiliser plots only, 3 weeks after transplanting (WAT). The crop was mainly rain-fed, with equal amounts of supplementary irrigation applied to all plots as required, using a watering can. The plots were kept weed free through hand hoeing. A repeat application of Malathion 50% E.C to control aphids was done at 5 WAT.

Table 1: Chemical composition of organic materials used for composting and as a soil amendment

	TSC	CM	TSC-CM
Total N (g/kg)	24.1	7.3	19.5
Total P (g/kg)	2.6	0.6	3.5
Total K (g/kg)	18.2	10.4	12.3
Total C (g/kg)	415	301	202
C/N	17.2	41.2	10.3
pH (H_2O)	5.9	6.8	8.7
Nicotine (mg/kg)	17 500	–	< 100

TSC, tobacco leaf scrap; CM, cattle manure; TSC-CM, tobacco leaf scrap and cattle manure compost

Biomass yield determination

Fresh oilseed rape leaves were harvested from a netplot, consisting of the four inner rows leaving two border rows on each side of the plot, at 3, 5, 7 and 9 WAT. Mature leaves were plucked from each plant, leaving three young leaves. Thereafter, the leaf samples were weighed

to determine fresh biomass. Sub-samples of the biomass were oven dried at 70 °C for 72 h, weighed and processed for subsequent nutrient content determination. Dry biomass yield was calculated by multiplying the fresh biomass yield with a moisture correction factor. The remaining fresh leaf biomass samples were used for organoleptic testing.

Plant tissue nutrient and nicotine determination

Tissue nutrient concentration (total N, P, K, Fe, Zn, Cu and Mn) was obtained from digestion of oven dry plant tissue sample using sulfuric acid–selenium powder mixture on a block digester at 330 °C.[35] This was followed by absorbance readings at 650 nm and 880 nm for total N and P, respectively, after blue colour development using a UV-vis single beam spectrophotometer. Distilled water was used to dilute aliquots of the sample digest and then K and micronutrient concentrations were read off an atomic absorption spectrophotometer.[35] Total uptake of N, P and K in kg/ha were separately calculated as:

Nutrient uptake (kg/ha) = concentration of the nutrient (%) *dry matter (kg/ha) / 100 **Equation 1**

Nicotine analysis was done at the Tobacco Research Board chemistry laboratory providing commercial analytical services. The procedure involved steam distillation of ground material under strongly basic conditions, followed by acidification of the distillate and ultraviolet absorbance reading at 236 mμ, 259 mμ and 282 mμ on a UV-vis spectrophotometer.[36]

Taste determination

Taste determination of fresh leaf biomass was done at each harvest, with the help of untrained panelists. The use of untrained panelists ensured a measure of general acceptance and not quality control as is the case with wine, tea and cheese tasting that require trained assessors.[12] The taste panel of 15 people (aged 18–65) consisted of MUAST employees as well as surrounding community members who consented to participate in the study. About 25–30 leaves were composited from each replicate treatment, chopped and boiled for 5 min in 1 L of water in separate pots. The samples were cooled and served in small plastic boxes, labelled using numbers and arranged randomly. Tasting was done one person at a time to minimise bias. A glass of water was given to each panelist to rinse their palate after tasting each sample.[12] Taste scoring was done on a scale of 1 to 4, with a score awarded for each tasted sample. The scores were recorded on a score sheet. A score of 1 represents 'not bitter', 2 – 'mildly bitter', 3 – 'bitter' and 4 – 'extremely bitter'.

Soil chemical analyses

Soil was randomly sub-sampled from the plots and composited at 3 and 9 WAT (first and last harvest) for soil pH, EC, total C, mineral N, extractable P and exchangeable K. Mineral N was measured from fresh soil samples while the rest of the assays were done on air-dry soil sub-samples. Soil pH and EC were read from a 1:2.5 w/v soil:deionised water supernatant suspension.[37] Mineral N was extracted from soil using potassium sulfate followed by colourimetric determination of ammonium-N and nitrate-N which were summed together to obtain total mineral-N.[35] Extractable P was determined according to Olsen's extraction method, while exchangeable K was determined using the ammonium acetate method at pH 5.8 followed by absorbance reading on an atomic absorption spectrophotometer.[35] Total C was analysed following the Walkley–Black method with external heating.[37] Soil nicotine content was determined as previously described for plant samples.

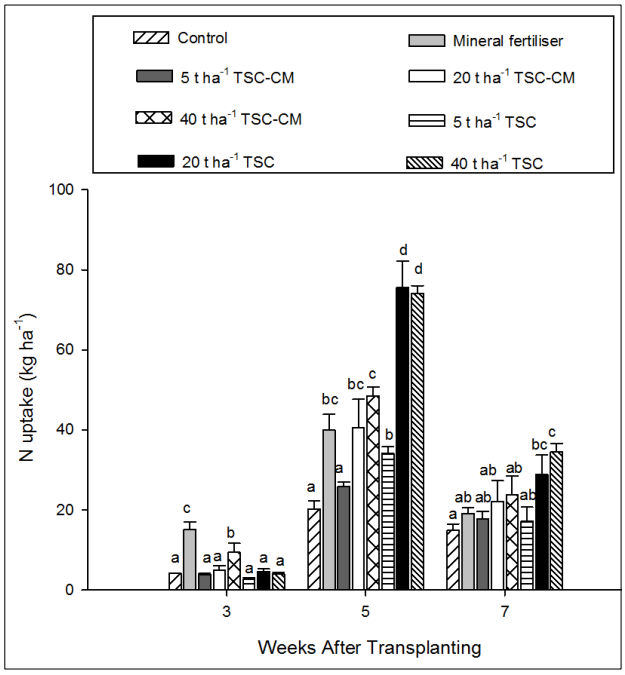
Statistical analysis

Dry biomass yield, macronutrient uptake, micronutrient and nicotine concentration, soil nutrient and nicotine content data, collected at each sampling interval, were subjected to a one-way analysis of variance (ANOVA) at $p<0.05$ using GenStat Discovery 14th edition.[38] Tukey's honestly significant difference was used to separate between statistically different means. Taste data were recorded, cleaned and analysed using SPSS v.21.[39] Taste data frequency rating for each treatment was derived by running descriptive statistical analysis in SPSS v.21 and frequency tables were generated. Thereafter, mean taste scores based on the rating of panelists for each treatment at each sampling interval were calculated and regressed against total N concentration in leaf biomass.[40] Pearson correlation coefficients were used to assess the strength of association.

Results and discussion

Oilseed rape macronutrient uptake

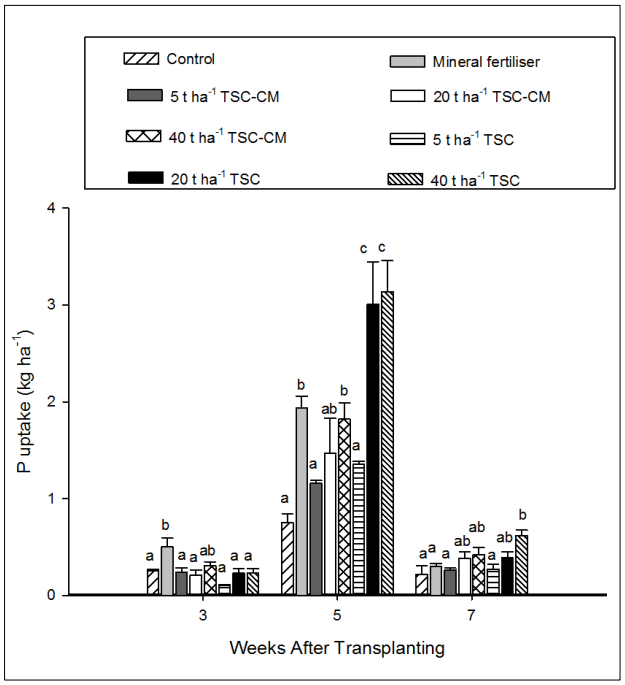
Nitrogen uptake at 3 WAT was higher ($p<0.05$) with mineral fertiliser and 40 t/ha TSC-CM than in the control and TSC treatments (Figure 1), while P uptake was more improved with the mineral fertiliser treatment (Figure 2). K uptake followed a similar trend to N uptake, although 20 t/ha TSC was comparable to mineral fertiliser and the TSC-CM treatments at 3 WAT (Figure 3). This highlighted enhanced nutrient uptake during early growth due to mineral fertiliser and TSC-CM application. While mineral fertiliser supplied nutrients in a ready form, the difference between TSC-CM and TSC could be attributed to ready mineralisation in TSC-CM on account of the narrower C/N ratio, indicative of more N per degradable C as a consequence of composting.[41] Additionally, depressed microbial activity associated with high nicotine content, as suggested by Adediran et al.[16], could have been the cause. At 5 WAT, N, P and K uptake was approximately three times more ($p<0.05$) than in the control with TSC application at 20 t/ha and 40 t/ha and between 1.5–2.0 times more than in mineral fertiliser and TSC-CM. This finding was not surprising as TSC could potentially supply more nutrients than mineral fertiliser and TSC-CM. For instance, at 40 t/ha TSC could potentially supply 964 kg N/ha relative to 121 and 780 kg N/ha from mineral fertiliser and 40 t/ha TSC-CM, respectively. At 7 WAT, only 40 t/ha TSC resulted in significantly higher N, P and K uptake relative to mineral fertiliser and TSC-CM application at all rates (Figures 1–3) pointing to the longevity of nutrient supply at the higher application rate. Mostly, nutrient uptake in 5 t/ha TSC-CM and TSC did not significantly vary from the control, indicating limited effectiveness of such low application rates of organic amendments. Observations of higher nutrient uptake and yield with increasing application rates of tobacco waste have been reported in other studies.[17,25]



TSC-CM, tobacco leaf scrap and cattle manure compost, TSC, uncomposted tobacco leaf scrap

Tukey's honestly significant different (p<0.05) means per sampling interval are represented by different alphabetical letters.

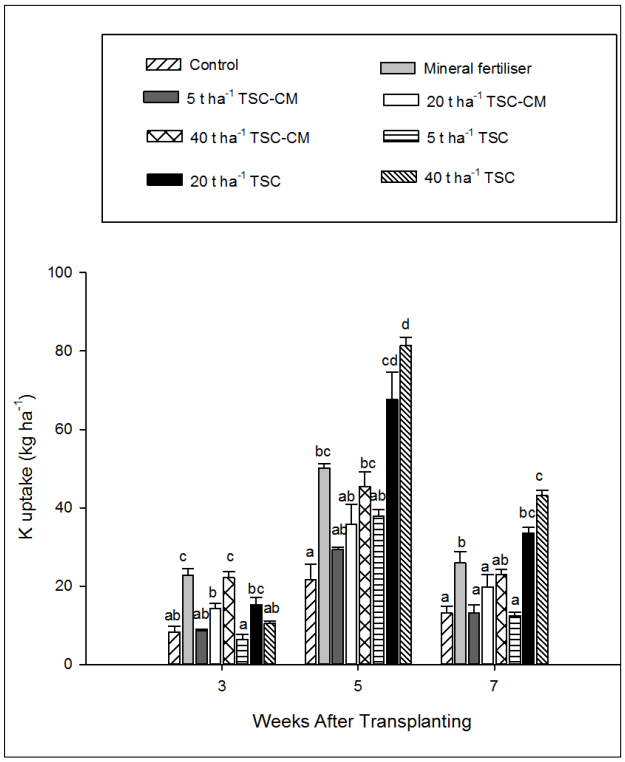
Figure 1: Oilseed rape (*Brassica napus* L.) leaf tissue N uptake (mean±s.e.).



TSC-CM, tobacco leaf scrap and cattle manure compost, TSC, uncomposted tobacco leaf scrap

Tukey's honestly significant different (p<0.05) means per sampling interval are represented by different alphabetical letters.

Figure 2: Oilseed rape (*Brassica napus* L.) leaf tissue P uptake (mean±s.e.).



TSC-CM, tobacco leaf scrap and cattle manure compost, TSC, uncomposted tobacco leaf scrap

Tukey's honestly significant different (p<0.05) means per sampling interval are represented by different alphabetical letters.

Figure 3: Oilseed rape (*Brassica napus* L.) leaf tissue K uptake (mean±s.e.).

Oilseed rape micronutrient content

Application of TSC-CM at 20 t/ha and 40 t/ha significantly ($p<0.05$) improved Mn at 3 and 7 WAT, Zn at 5 WAT and Cu across all harvests, relative to control and mineral fertiliser (Table 2). Fe, Mn and Cu were improved by both 20 t/ha and 40 t/ha TSC relative to control and mineral fertiliser at selected sampling intervals (Table 2). Unlike macronutrients, micronutrient concentration, for example Cu and Fe, was also improved by 5 t/ha application rate at certain sampling intervals. Micronutrient concentration has increasingly been reported to be low in soils on most smallholder farms resulting in calls for supplementation through foliar applications or adoption of micronutrient-enriched mineral fertilisers in order to meet human dietary needs.[42] Therefore, TSC-CM and TSC offer a cheaper source of micronutrients on smallholder farms as observed in this study and in other studies.[27]

Nicotine content and toxicity risk in oilseed rape

Nicotine content in all oilseed rape leaf sub-samples was below detectable levels (<0.01% of dry leaf matter) indicating assimilation in low levels. The nicotine concentration was thus below 500 mg/kg toxicity threshold according to European Union Regulations.[43] This finding was expected, especially from TSC-CM which had undergone prior composting.[16] With regard to TSC, rapid degradation of nicotine has been reported to occur between a few hours to a few days in soil in other studies,[32,33] thus explaining its low assimilation by the plants. Subsequently, risk of nicotine toxicity from consumption of the oilseed rape fertilised with TSC was minimal.

Oilseed rape yield response to fertilisation

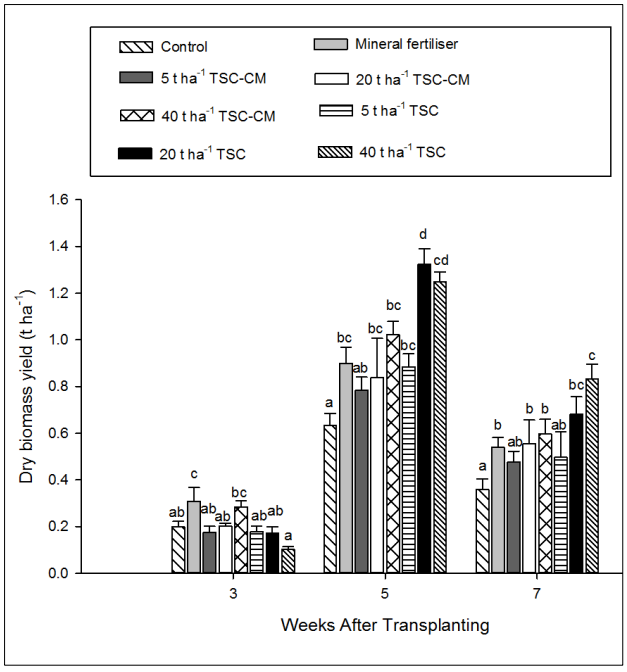
At 3 WAT, only mineral fertiliser and 40 t/ha TSC-CM treated plots had dry biomass yield that was higher ($p<0.05$) than that of the control, while those of the other treatments were not different (Figure 4). This finding was attributed to availability of nutrients in a ready form from mineral fertiliser, while only 40 t/ha application of TSC-CM ensured supply of nutrients in adequate amounts. On the other hand, although nutrient levels in TSC were comparatively higher (Table 1), there was wilting of seedlings in the 40 t/ha TSC treatment during the first week, resulting in gap filling being done in the second week. This was indicative of unfavourable conditions when high doses of TSC were applied. This observation is similar to that of Seckar et al.[33] who reported inhibited early growth of lettuce head with high rates of tobacco waste amendment ascribed to high levels of nicotine in soil. In another study, Adediran et al.[16] attributed poor seed germination of lettuce to high EC of tobacco waste and cabbage compost. Nonetheless, the inhibitory conditions were shown to subside within a week as successful gap filling was achieved in the subsequent week. This is consistent with the observations of Nota et al.[32] and Seckar et al.[33] who put forward that nicotine degradation in soils takes a few days, after which the risk of toxicity is extremely low. Consequently, this could have been the case in this study. Furthermore, the effects were only observable with the 40 t/ha application rate of TSC pointing to a dose-dependent effect.[44] Although Delibacak and Ongun[15] recommended a 1-month waiting period before planting or sowing activities when soil is amended with uncomposted tobacco waste, the present study indicated that a shorter waiting period of at least 1 week may be adequate. A shorter waiting period may reduce opportunities for nutrient loss, thus fostering optimum nutrient utilisation. At 5 WAT, 40 t/ha TSC resulted in 116%, 85% and 30% more biomass yield than control, mineral fertiliser and 40 t/ha TSC-CM, respectively (Figure 4). Further, at 7 WAT, more biomass yield was realised from 20 and 40 t/ha TSC relative to control, mineral fertiliser and TSC-CM (Figure 4). At 9 WAT, TSC at 20 t/ha and 40 t/ha yielded approximately 0.25 t/ha and 0.34 t/ha dry biomass, respectively (not shown on the figure), while the rest of the treatments did not produce any harvestable biomass, due to severe damage of plants by aphids (*Aphidoidea* spp.). Resultantly, cumulative biomass yield was significantly higher in 20 and 40 t/ha TSC treated plots relative to those of other treatments including mineral fertiliser and TSC-CM amended treatments (Table 3). This was attributed to supply of more nutrients from TSC relative to TSC-CM and mineral fertiliser as well as the additional biomass harvested at 9 WAT.

Table 2: Micronutrient concentration in oilseed rape (*Brassica napus* L.) leaf tissue

	Weeks after transplanting											
	3	5	7	3	5	7	3	5	7	3	5	7
	Fe (mg/kg)			Zn (mg/kg)			Mn (mg/kg)			Cu (mg/kg)		
Control	356.2^{a}	243.0^{a}	174.3^{a}	72.1^{a}	69.0^{a}	61.5^{a}	93.0^{a}	45.1^{a}	27.3^{a}	7.1^{a}	3.3^{a}	7.1^{a}
Mineral fertiliser	521.3^{a}	295.4^{ab}	198.4^{ab}	86.3^{ab}	78.4^{ab}	75.8^{ab}	104.6^{ab}	42.7^{a}	27.9^{a}	7.6^{a}	4.6^{a}	10.1^{ab}
5 t/ha TSC-CM	681.6^{ab}	203.6^{a}	164.2^{a}	139.6^{b}	132.7^{c}	74.6^{ab}	98.3^{a}	42.3^{a}	31.8^{ab}	18.8^{b}	18^{bcd}	15.4^{bc}
20 t/ha TSC-CM	845.4^{ab}	256.9^{ab}	251.7^{bc}	147.6^{b}	136.5^{c}	83.5^{b}	132.0^{cd}	53.2^{a}	41.9^{bc}	23.5^{b}	21.5^{cd}	20.9^{cd}
40 t/ha TSC-CM	910.8^{b}	342.1^{bc}	261.6^{bc}	204.4^{c}	179.3^{d}	85.6^{b}	144.8^{cde}	65.8^{b}	49.9^{bc}	26.6^{b}	23.4^{d}	20.8^{cd}
5 t/ha TSC	1559.1^{c}	292.0^{ab}	173.7^{a}	91.8^{ab}	71.4^{ab}	73.5^{ab}	110.8^{ab}	41.9^{a}	34.8^{ab}	8.5^{a}	3.9^{a}	6.1^{a}
20 t/ha TSC	1579.4^{c}	331.2^{bc}	224.6^{bc}	103.2^{ab}	85.1^{ab}	92.8^{bc}	111.5^{ab}	50.2^{a}	45.6^{bc}	8.6^{a}	8.3^{ab}	14.1^{bc}
40 t/ha TSC	1154.4^{bc}	392.8^{bc}	221.8^{bc}	131.2^{b}	100.7^{b}	107.6^{c}	150.7^{de}	74.9^{bc}	65.2^{d}	9.0^{a}	12.1^{bc}	17.9^{c}

Means followed by the same lowercase alphabetical letters at each harvest are not significantly different at p<0.05 according to Tukey's honestly significant difference test.

TSC, tobacco leaf scrap; CM, cattle manure; TSC-CM, tobacco leaf scrap and cattle manure compost



TSC-CM, tobacco leaf scrap and cattle manure compost, TSC, uncomposted tobacco leaf scrap

Tukey's honestly significant different (p<0.05) means per sampling interval are represented by different alphabetical letters.

Figure 4: Oilseed rape (*Brassica napus* L.) dry biomass yield (mean±s.e.).

Table 3: Cumulative dry mass yield of oilseed rape (*Brassica napus* L.) 9 weeks after transplanting

Treatment	Dry biomass yield (t/ha)
Control	1.19^{a}
Mineral fertiliser	1.75^{b}
5 t/ha TSC-CM	1.43^{ab}
20 t/ha TSC-CM	1.61^{b}
40 t/ha TSC-CM	1.91^{b}
5 t/ha TSC	1.56^{ab}
20 t/ha TSC	2.43^{c}
40 t/ha TSC	2.52^{c}

Means followed by the same lowercase alphabetical letters are not significantly different at p<0.05 according to Tukey's honestly significant difference test.

TSC, tobacco leaf scrap; CM, cattle manure; TSC-CM, tobacco leaf scrap and cattle manure compost

Oilseed rape organoleptic testing

The control was rated between 'not bitter' and 'mildly bitter' across the three sampling intervals, reflected by corresponding mean taste scores (Table 4). Extremely bitter taste of oilseed rape was observed with mineral fertiliser and TSC at 20 t/ha and 40 t/ha at 3 WAT and decreased with time (Table 4). On the other hand, TSC-CM at all rates and 5 t/ha TSC ranged between 'mildly bitter' and 'bitter' at 3 WAT, diminishing with time (Table 4). There was a positive but weak correlation between total N content and mean taste scores across all harvest periods (Figure 5), while no correlations were done with nicotine concentration owing to non-detectable levels. The intense bitterness of oilseed rape in TSC treatments at 20 t/ha and 40 t/ha confirmed the existing but undocumented concerns in the study area with TSC use, although it was not limited to TSC treatments – it also occurred with mineral fertilisation. While the bitterness with mineral fertilisation could be attributed to nitrogen and nitrate content in *Brassica juncea* leaves, as reported by Maereka et al.[12], a weak correlation between taste and leaf N observed in this study indicated that nitrate could have had a more pronounced effect than total N. On the other hand, nitrate effects coupled with possible assimilation of by-products of nicotine breakdown could have been responsible for the bitter taste in TSC amended treatments,[45] especially as bitterness was more intense during the early growth stages which coincided with periods of high nicotine degradation. Nevertheless, these effects subsided with time, thus pointing to an opportunity for utilisation.

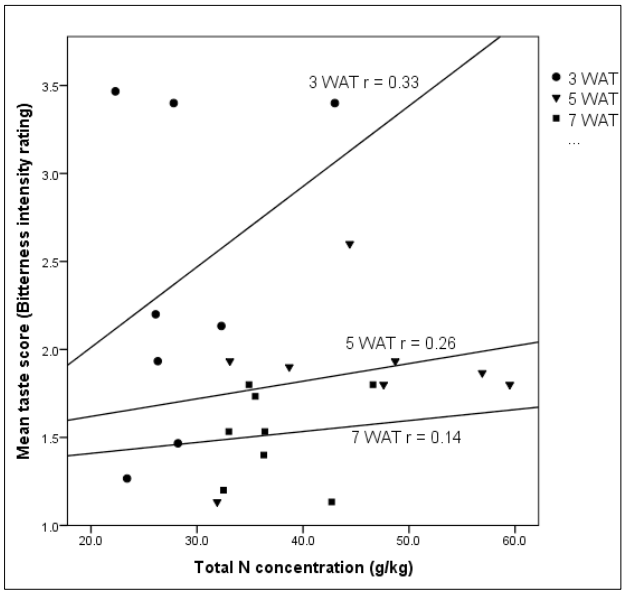
Table 4: Frequency of panelists' taste perception of oilseed rape (*Brassica napus* L.) leaf taste and mean taste scores

WAT	Treatment	Not bitter	Mildly bitter	Bitter	Extremely bitter	Mean taste score
3 WAT	Control	73.3	26.7	0.0	0.0	1.3
	Mineral fertiliser	0.0	13.3	33.3	53.4	3.4
	5 t/ha TSC-CM	40.0	26.7	33.3	0.0	1.9
	20 t/ha TSC-CM	60.0	33.3	6.7	0.0	1.5
	40 t/ha TSC-CM	6.7	73.3	20.0	0.0	2.1
	5 t/ha TSC	13.3	53.4	33.3	0.0	2.2
	20 t/ha TSC	0.0	0.0	60	40.0	3.4
	40 t/ha TSC	0.0	0.0	53.3	46.7	3.5
5 WAT	Control	86.7	13.3	0.0	0.0	1.1
	Mineral fertiliser	0.0	40.0	60.0	0.0	2.6
	5 t/ha TSC-CM	33.3	40.0	26.7	0.0	1.9
	20 t/ha TSC-CM	33.3	53.3	13.3	0.0	1.8
	40 t/ha TSC-CM	20.0	66.7	13.3	0.0	1.9
	5 t/ha TSC	26.7	60.0	13.3	0.0	1.9
	20 t/ha TSC	6.7	80.0	13.3	0.0	1.8
	40 t/ha TSC	0.0	33.3	60.0	6.7	2.0
7 WAT	Control	80.0	20.0	0.0	0.0	1.2
	Mineral fertiliser	40.0	46.7	13.3	0.0	1.7
	5 t/ha TSC-CM	53.7	47.3	0.0	0.0	1.5
	20 t/ha TSC-CM	60.0	40.0	0.0	0.0	1.4
	40 t/ha TSC-CM	86.7	13.3	0.0	0.0	1.1
	5 t/ha TSC	40.0	60.0	0.0	0.0	1.5
	20 t/ha TSC	60.0	26.7	13.3	0.0	1.8
	40 t/ha TSC	33.3	53.4	13.3	0.0	1.8

The values are valid percentage of respondents (n=15).

A taste score of 1 represents 'not bitter', 2 – 'mildly bitter', 3 – 'bitter' and 4 – 'extremely bitter'.

TSC, tobacco leaf scrap; CM, cattle manure; TSC-CM, tobacco leaf scrap and cattle manure compost; WAT, weeks after transplanting



WAT, weeks after transplanting

Figure 5: Relationship between total N content and mean taste scores of oilseed rape (*Brassica napus* L.).

Selected soil chemical properties during growth of oilseed rape

Soil pH ranged from 6.0 to 7.1 and did not significantly ($p>0.05$) vary across treatments at 3 WAT, although at 9 WAT it was significantly ($p<0.05$) higher with 20 and 40 t/ha TSC-CM (Table 5). The alkaline nature of TSC-CM (Table 1) most likely contributed to the increase in soil pH and this was desirable as the sandy soils in the study area are prone to acidity and increased risk of induced nutrient deficiencies.[46] Consequently, TSC-CM application can be recommended to curb these problems. At 3 WAT, mineral fertiliser and TSC-CM produced higher EC than control and TSC applications, while at 9 WAT, both TSC-CM and TSC at 20 t/ha and 40 t/ha produced higher EC than both the control and mineral fertiliser did (Table 5). This indicated higher cationic and anionic concentrations, mostly likely mineralised from the organic amendments. Bertran et al.[47] and Loper at al.[48] attributed a similar increase of EC to cation release during organic material mineralisation. The increase in EC was variable and although 40 t/ha TSC resulted in 12 times higher EC than control, it was still below the critical limit of 4.0 dS/m and thus did not pose any detrimental salt effects.[49] Soil organic carbon was significantly ($p<0.05$) higher with application of both TSC-CM (except at 5 t/ha) and TSC than the control and mineral fertiliser at 3 WAT (Table 5). Also at 9 WAT, both 20 t/ha and 40 t/ha application rates for TSC-CM and TSC resulted in significantly higher SOC over control and mineral fertiliser. This was by a margin of 85–169% with TSC-CM, while a wider margin of 137–229% was noted with TSC application. Higher SOC content with TSC at 9 WAT was in accordance with the higher C content in TSC due to availability of undegraded substrate. Additionally, carbon shielding associated with possible aggregate binding due to the release of phenolic substances during decomposition of TSC could have contributed to the higher SOC.[31] On the other hand, the significantly ($p<0.05$) higher mineral N after addition of 20 t/ha and 40 t/ha TSC-CM and TSC relative to the control treatment (Table 5) was ascribed to the supply of nutrients from the organic amendments. Mineralisation of the organic materials also explained the observed high levels of extractable P and exchangeable K (Table 5). Generally, higher nutrient contents were observed in TSC-CM at 3 WAT than TSC owing to the narrower C/N ratio in the compost which favoured faster mineralisation.[41] This was also confirmed by the nutrient uptake observations and pointed to more ready mineralisation of TSC-CM than TSC, while conversely, TSC presented an opportunity for prolonged supply of nutrients. Nicotine content in soil was below the 100 mg/kg detection level in all treatments across sampling periods (results not shown). As such, the risk of nicotine toxicity with TSC application was shown not to persist beyond 3 weeks. This observation attests to observations of swift degradation of nicotine in soil, as reported in other studies.[32,33]

Table 5: Selected soil chemical properties during growth of oilseed rape (*Brassica napus* L.)

Treatment	Weeks after transplanting											
	3	9	3	9	3	9	3	9	3	9	3	9
	pH H_2O		EC (μS/cm)		Total C (%)		Mineral N (mg/g)		Available P (mg/kg)		Exchangeable K (cmol/kg)	
Control	6.52^{a}	6.59^{a}	62.1^{ab}	40.3^{a}	0.49^{a}	0.47^{a}	3.5^{a}	<1^{a}	20.8^{a}	19.6^{a}	0.21^{a}	0.19^{a}
Mineral fertiliser	6.01^{a}	6.30^{a}	112.7^{b}	52.1^{a}	0.55^{a}	0.49^{a}	60.0^{c}	15^{a}	102.1^{c}	67.9^{b}	0.73^{c}	0.56^{b}
5 t/ha TSC-CM	6.79^{a}	6.91^{ab}	115.0^{b}	48.0^{a}	0.52^{a}	0.66^{a}	10^{ab}	7.5^{a}	21.5^{a}	20.8^{a}	0.41^{b}	0.20^{a}
20 t/ha TSC-CM	6.92^{a}	7.08^{b}	233.0^{c}	172.5^{c}	1.86^{c}	1.34^{bc}	28.0^{b}	17.5^{a}	79.0^{b}	36.1^{a}	0.85^{cd}	0.23^{a}
40 t/ha TSC-CM	6.97^{a}	7.05^{b}	322.0^{d}	168.0^{c}	1.89^{c}	1.96^{de}	40.0^{bc}	20.0^{a}	90.8^{bc}	136.3^{d}	0.91^{d}	0.24^{a}
5 t/ha TSC	6.44^{a}	6.17^{a}	48.3^{a}	40.5^{a}	1.28^{b}	0.93^{ab}	7.5^{a}	30.0^{a}	32.7^{a}	38.2^{a}	0.13^{a}	0.23^{a}
20 t/ha TSC	6.41^{a}	6.60^{a}	77.0^{ab}	128.0^{b}	1.32^{bc}	1.66^{cd}	20.0^{b}	40.0^{a}	33.3^{a}	91.0^{c}	0.21^{a}	0.45^{b}
40 t/ha TSC	6.75^{a}	6.62^{a}	98.5^{ab}	505.0^{d}	1.44^{bc}	2.31^{e}	25.0^{b}	140.0^{b}	38.4^{a}	192.4^{e}	0.21^{a}	1.47^{c}

Means followed by the same lowercase alphabetical letters at each harvest are not significantly different at $p<0.05$ *according to Tukey's honestly significant difference test.*

TSC, tobacco leaf scrap; CM, cattle manure; TSC-CM, tobacco leaf scrap and cattle manure compost

Conclusion

The readiness of supply of nutrients by TSC-CM was highlighted at an application rate of 40 t/ha, and pointed to its importance in the promotion of early crop growth. On the other hand, higher nutrient supply and longevity of supply were realised with TSC application at 20 t/ha and 40 t/ha. In addition, while soil pH was improved with TSC-CM application, both TSC-CM and TSC improved soil nutrient levels including SOC, which is key towards organic matter build-up. Moreover, both TSC-CM and TSC were comparable in improving micronutrient concentration in oilseed rape relative to control and mineral fertiliser. Interestingly, there was low risk of nicotine toxicity in oilseed rape and soil with TSC application from 3 WAT. However, owing to the intensely bitter taste of oilseed rape in TSC-amended treatments, application of TSC-CM at 40 t/ha offers a trade-off of yield increase, nutrient uptake and oilseed rape taste and may, therefore, be more acceptable as an organic nutrient source relative to TSC. Notwithstanding, if composting is not feasible, TSC at 20 t/ha may be recommended, although the first harvest may have to be discarded. Thus, the viability of TSC-CM and TSC as practical solutions for nutrient supply and sustainable soil management were asserted. For further study, there is need for evaluation of performance of the organic materials based on the same N rate, evaluation of waiting periods after incorporation of TSC so as to enhance nutrient uptake during early growth stage as well as long-term application for sustainability.

Acknowledgements

We thank the Organisation for Women in Science for the Developing World for the PhD fellowship awarded to N.D. and the International Foundation for Science for a research grant (grant C/5764-1). We acknowledge the analytical services provided by the chemistry laboratory at the Tobacco Research Board.

Competing interests

We declare that there are no competing interests.

Authors' contributions

N.D.: Conceptualisation, methodology, data collection, sample analysis, validation, data curation, writing – the initial draft, writing – revisions, project leadership, project management, funding acquisition. R.Z.: Conceptualisation, methodology, student supervision, project management. C.P.: Methodology, data collection, sample analysis, writing – revisions. M.W.: Conceptualisation, methodology, student supervision, writing – revisions. P.M.: Conceptualisation, methodology, student supervision, writing – revisions.

References

1. Mandiriza-Mukwirimba G, Kritzinger Q, Aveling T. A survey of brassica vegetable smallholder farmers in the Gauteng and Limpopo provinces of South Africa. J Agric Rural Dev Trop Subtrop. 2016;117:35–44.
2. Maseko I, Mabhaudhi T, Tesfay S, Araya HT, Fezzehazion M, Du Plooy CP. African leafy vegetables: A review of status, production and utilization in South Africa. Sustainability. 2018;10(1), Art. #16. https://doi.org/10.3390/su10010016
3. Mango N, Makate C, Mapemba L, Sopo M. The role of crop diversification in improving household food security in central Malawi. Agric Food Secur. 2018;7(1), Art. #7. https://doi.org/10.1186/s40066-018-0160-x
4. Acikgoz FE. Mineral, vitamin C and crude protein contents in kale (*Brassica oleraceae* var. *acephala*) at different harvesting stages. Afr J Biotechnol. 2011;10:17170–17174. http://dx.doi.org/10.5897/AJB11.2830
5. Ganya S, Svotwa E, Katsaruware RD. Performance of two rape (*Brassica napus*) cultivars under different fertilizer management levels in the smallholder sector of Zimbabwe. Int J Agron. 2018;2018, Art. #2351204, 7 pages. https://doi.org/10.1155/2018/2351204
6. SNV (Netherlands Development Organisation). Smallholder horticultural production and business trainer's manual. Harare: SNV Zimbabwe; 2016. Available from: http://www.snv.org/public/cms/sites/default/files/explore/download/rarp_2016-horticulture-trainers-manual.pdf
7. Mungai LM, Snapp S, Messina JP, Chikowo R, Smith A, Anders E, et al. Smallholder farms and the potential for sustainable intensification. Front Plant Sci. 2016;7:1–17. https://doi.org/10.3389/fpls.2016.01720
8. Musara C, Chitamba J. Growth rate and yield of *Brassica napus* in response to *Acacia angustissima* leaf biomass application. J Anim Plant Sci. 2015;25:510–518. http://www.thejaps.org.pk/docs/v-25-02/26.pdf
9. Muchecheti BF, Madakadze IC. Yield and nitrogen recovery of rape (*Brassica napus* L.) in response to application of leguminous leaf litter and supplemental inorganic nitrogen. Exp Agric. 2016;52:518–536. https://doi.org/10.1017/S0014479715000228
10. Cedrez BC, Chamberlin J, Guo Z, Hijmans RJ. Spatial variation in fertilizer prices in sub-Saharan Africa. PLoS ONE. 2020;15:1–20. http://dx.doi.org/10.1371/journal.pone.0227764
11. Khan MN, Mobin M, Abbas ZK, Alamri SA. Fertilizers and their contaminants in soils, surface and groundwater. In: DellaSala DA, Goldstein MI, editors. The encyclopedia of the Anthropocene. Oxford: Elsevier Inc.; 2018. p. 225–240. https://doi.org/10.1016/B978-0-12-809665-9.09888-8
12. Maereka EK, Madakadze RM, Mashingaidze AB, Kageler S, Nyakanda C. Effect of nitrogen fertilization and timing of harvesting on leaf nitrate content and taste in mustard rape (*Brassica juncea* L. *Czern*). J Food Agric Environ. 2007;5:288–293.
13. Mehdizadeh M, Darbandi EI, Naseri-rad H, Tobeh A. Growth and yield of tomato (*Lycopersicon esculentum* Mill.) as influenced by different organic fertilizers. Int J Agron Plant Prod. 2013;4:734–738.
14. Materechera SA. Utilization and management practices of animal manure for replenishing soil fertility among smallscale crop farmers in semi-arid farming districts of the North West Province, South Africa. Nutr Cycl Agroecosystems. 2010;87:415–428. https://doi.org/10.1007/s10705-010-9347-7
15. Delibacak S, Ongun AR. Influence of composted tobacco waste and farmyard manure applications on the yield and nutrient composition of lettuce. Eurasian J Soil Sci. 2016;5:132–138. https://doi.org/10.18393/ejss.2016.2.132-138
16. Adediran JA, Mnkeni PNS, Mafu NC, Muyima NYO. Changes in chemical properties and temperature during the composting of tobacco waste with other organic materials, and effects of resulting composts on lettuce (*Lactuca sativa* L.) and spinach (*Spinacea oleracea* L.). Biol Agriulture Hortic. 2004;22:101–119. https://doi.org/10.1080/01448765.2004.9754994
17. Chaturvedi S, Upreti DK, Tandon DK, Sharma A, Dixit A. Bio-waste from tobacco industry as tailored organic fertilizer for improving yields and nutritional values of tomato crop. J Environ Biol. 2008;29:759–763.
18. Okur N, Hüsnü H, Lu K, Okur B, Delibacak S. Organic amendment based on tobacco waste compost and farmyard manure: Influence on soil biological properties and butter-head lettuce yield. Turksh J Agric For. 2008;32:91–99.
19. Gülser C, Candemir F. Changes in penetration resistance of a clay field with organic waste applications. Eurasian J Soil Sci. 2012;1:16–21.
20. Kayikc¸iog˘lu HH, Okur N. Evolution of enzyme activities during composting of tobacco waste. Waste Manag Res. 2011;29:1124–1133. https://doi.org/10.1177/0734242x10392813
21. Prowse M. A history of tobacco production and marketing in Malawi, 1890–2010. J East Afr Stud. 2013;7:691–712. https://doi.org/10.1080/17531055.2013.805077
22. Scoones I, Mavedzenge B, Sukume C, Murimbarimba F. Tobacco, contract farming, and agrarian change in Zimbabwe. J Agrar Chang. 2017:1–21. https://doi.org/10.1111/joac.12210
23. Masvongo J, Mutambara J, Zvinavashe A. Viability of tobacco production under smallholder farming sector in Mount Darwin District, Zimbabwe. J Dev Agric Econ. 2013;5:295–301. https://doi.org/10.5897/JDAE12.128
24. Hu T, Lee AH. Tobacco control and tobacco farming in African countries. J Public Health Policy. 2015;36:1–9. https://doi.org/10.1057/jphp.2014.47
25. Cercioglu M, Okur B, Delibacak S, Ongun AR. Effects of tobacco waste and farmyard manure on soil properties and yield of lettuce (*Lactuca sativa* L. var. *Capitata*). Commun Soil Sci Plant Anal. 2012;43:875–886. https://doi.org/10.1080/00103624.2012.653023
26. Wang SN, Liu Z, Fang HZ, Meng J, Xu P. Characterization of environmentally friendly nicotine degradation by *Pseudomonas putida* biotype A strain S16. Microbiology. 2007;153:1556–1565. https://doi.org/10.1099/mic.0.2006/005223-0

27. Adediran JA, De Baets N, Mnkeni PNS, Kiekens L, Muyima NYO, Thys A. Organic waste materials for soil fertility improvement in the Border Region of the Eastern Cape, South Africa. Biol Agric Hortic. 2003;20:283–300. https://doi.org/10.1080/01448765.2003.9754974

28. Dunjana N, Zengeni R, Muchaonyerwa P, Wuta M. Typological characterisation of farms in a smallholder food-cash crop production system in Zimbabwe – opportunities for livelihood sustainability. J Agric Rural Dev Trop Subtrop. 2018;119:11–22.

29. Gulser C, Demir Z, Serkan I. Changes in some soil properties at different incubation periods after tobacco waste application. J Environ Biol. 2010;31:671–674.

30. Akhtar S, Shakeel S, Hamid A, Ahmad SR. Suitability of tobacco dust for agricultural purposes. J Biodivers Environ Sci. 2016;9:102–113.

31. Shakeel, S. Consideration of tobacco dust as organic amendment for soil: A soil & waste management strategy. Earth Sci. 2014,3:117-121. https://doi.org/10.11648/j.earth.20140305.11

32. Nota G, Naviglio D, Ugliano M, Romano R. Determination of nicotine in the soil mixed with tobacco powder as fertilizer. Anal Lett. 2008;33:265–275. https://doi.org/10.1080/00032710008543051

33. Seckar JA, Stavanja MS, Harp PR, Yi Y, Garner CD, Doi J. Environmental fate and effects of nicotine released during cigarette production. Environ Toxicol Chem. 2008;27:1505–1514. https://doi.org/10.1897/07-284.1

34. Food and Agriculture Organisation(FAO). World reference base for soil resources 2014: International soil classification system for naming soils and creating legends for soil maps. Rome: FAO; 2015. Available from: http://www.fao.org/soils-portal/soil-survey/soil-classification/world-reference-base/en/

35. Okalebo JR, Gathua KW, Woomer PL. Laboratory methods of soil and plant analysis: A working manual. 2nd ed. Nairobi: TSBF-CIAT; SACRED Africa; 2002.

36. Willits C, Swain ML, Connelly JA, Brice BA. Spectrophotometric determination of nicotine. Anal Chem. 1950;22:430–433. https://doi.org/10.1021/ac60039a013

37. Anderson JM, Ingram JSI. Tropical soil biology and fertility (TSBF). A hand book of methods. 2nd ed. Wallingford, UK: CAB International; 1993.

38. VSN International. GenStat for Windows. 14th ed. Hemel Hempstead, UK.: VSN International; 2013.

39. IBM SPSS. Statistics for Windows. Armonk, NY: IBM Corp.; 2012.

40. Tordoff MG, Sandell MA. Vegetable bitterness is related to calcium content. Appetite. 2009;52:498–504. https://doi.org/10.1016/j.appet.2009.01.002

41. Bernal MP, Alburquerque JA, Moral R. Composting of animal manures and chemical criteria for compost maturity assessment. A review. Bioresour Technol. 2009;100:5444–5453. http://dx.doi.org/10.1016/j.biortech.2008.11.027

42. Manzeke GM, Mtambanengwe F, Nezomba H, Mapfumo P. Zinc fertilization influence on maize productivity and grain nutritional quality under integrated soil fertility management in Zimbabwe. Field Crops Res. 2014;166:128–136. http://dx.doi.org/10.1016/j.fcr.2014.05.019

43. Civilini M, Domenis C, Sebastianutto N, de Bertoldi M. Nicotine decontamination of tobacco agro-industrial waste and its degradation by micro-organisms. Waste Manag Res. 1997;15:349–358. https://doi.org/10.1177/0734242X9701500403

44. Selmar D, Engelhardt UH, Hänsel S, Thräne C, Selmar D, Engelhardt UH, et al. Nicotine uptake by peppermint plants as a possible source of nicotine in plant-derived products. Agron Sustain Dev. 2015;35:1185–1190. https://doi.org/10.1007/s13593-015-0298-x

45. Yu H, Tang H, Xu P. Green strategy from waste to value-added-chemical production: Efficient biosynthesis of 6-hydroxy-3-succinoyl-pyridine by an engineered biocatalyst. Sci Rep. 2014;4:1–8. http://10.1038/srep05397

46. Soropa G, Nyamangara J, Nyakatawa EZ. Nutrient status of sandy soils in smallholder areas of Zimbabwe and the need to develop site-specific fertiliser recommendations for sustainable crop intensification. S AfrJ Plant Soil. 2018;1–3. https://doi.org/10.1080/02571862.2018.1517901

47. Bertran E, Sort X, Soliva M, Trillas I. Composting winery waste: Sludges and grape stalks. Bioresour Technol. 2004;95:203–208. https://doi.org/10.1016/j.biortech.2003.07.012

48. Loper S, Shober AL, Wiese C, Denny GC, Stanley CD, Gilman EF, et al. Organic soil amendment and tillage affect soil quality and plant performance in simulated residential landscapes. HortScience. 2010;45:1522–1528. https://doi.org/10.21273/HORTSCI.45.10.1522

49. Angelova VR, Akova VI, Artinova NS, Ivanov KI. The effect of organic amendments on soil chemical characteristics. Bulg J Org Agric Acad. 2013;19:958–971.

A portfolio perspective of rural livelihoods in Bushbuckridge, South Africa

AUTHORS:
Fatima H. Ragie[1]
David W. Olivier[2]
Lori M. Hunter[3,4]
Barend F.N. Erasmus[2]*
Coleen Vogel[2]
Mark Collinson[4,5]
Wayne Twine[1]

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]Institute of Behavioral Science, CU Population Center, Department of Sociology, University of Colorado Boulder, Boulder, Colorado, USA
[4]MRC/Wits-Rural Public Health and Health Transitions Unit (Agincourt), School of Public Health, University of the Witwatersrand, Johannesburg, South Africa
[5]DSI/MRC South African Population Research Infrastructure Network (SAPRIN), University of Johannesburg, Johannesburg, South Africa
Present address: Faculty of Natural and Agricultural Sciences, University of Pretoria, Pretoria, South Africa

CORRESPONDENCE TO:
David Olivier

EMAIL:
david.olivier@wits.ac.za

DATES:
Received: 25 Oct. 2019
Revised: 14 Feb. 2020
Accepted: 28 Apr. 2020
Published: 29 Sep. 2020

HOW TO CITE:
Ragie FH, Olivier DW, Hunter ML, Erasmus BFN, Vogel C, Collinson M, et al. A portfolio perspective of rural livelihoods in Bushbuckridge, South Africa. S Afr J Sci. 2020;116(9/10), Art. #7522, 8 pages. https://doi.org/10.17159/sajs.2020/7522

ARTICLE INCLUDES:
☒ Peer review
☐ Supplementary material

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

Land-based income streams, which include the consumption and selling of crops, livestock and environmental products, are inherent in rural households' livelihoods. However, the off-farm cash income stream – primarily composed of migrant labour remittances, social grants, and savings and loans – is increasing in importance in many regions. This case study of 590 households from Bushbuckridge, South Africa, analyses the economic value of each of these income streams at three points: what enters the household, what is used and what is sold. Two important findings emerge. First, dependence on off-farm cash incomes is far higher than previously suggested by case studies in the area and the benefits of employment accrue to those already better educated and wealthier. This suggests that shifts in off-farm opportunities will exacerbate already deep inequalities. Second, while environmental products and crops are important for direct use, they generate insignificant cash incomes from sales. This suggests a weakening of the direct links between the local ecosystem and this society, challenging traditional notions of African rurality being intrinsically land based.

Significance:

Off-farm incomes such as wage labour, remittances and social grants are almost the sole source of cash for households in the study area. Even when including non-monetary incomes such as harvested produce, foraged goods and livestock products, off-farm incomes still represent the overwhelmingly largest proportion of overall household income value. This highlights the fact that South African rural economies are not consistently or primarily land based, and indicates the necessity of rural development strategies that facilitate participation in local cash economies. Otherwise, such efforts will be unable to yield broad benefits and will, instead, simply enrich those who are already better off.

Introduction

Household incomes in rural areas have been historically land-based, involving activities like cropping, livestock farming and resource harvesting, and were thus directly dependent on local ecosystem services.[1] However, a considerable body of scholarship records the deagrarianisation of rural economies resulting from increasingly accessible industries, markets and infrastructure.[2-4] In South Africa, increases in off-farm cash incomes are also driven by unprecedented access to information and communication technologies, as well as higher levels of education and modern attitudes in young adults from rural areas who are both pressured and aspire to exit land-based livelihoods.[5]

While this shift presents opportunities for some rural South African households, growing dependence on cash incomes may marginalise those individuals who are less likely to be able to afford to buy goods or pay for basic services, nor have the financial or human capital to engage in migrant labour.[5,6] Unchecked economic development in rural areas may therefore create greater inequality, while only benefitting the few households that are already economically well positioned.

A growing dependence on off-farm cash incomes has been noted in the study area[7,8], indicating weakening feedback loops between a rural society and its surrounding ecology[1]. The wealth of some rural households may protect their livelihood against a depleted local environment, but for many, having a range of land-based income streams provides basic needs such as food security and resilience against seasonal and cyclical stresses, as well as against sudden shocks, such as the death of a household member.[9,10]

Environmental products in particular are generally freely available and are therefore a useful safety net from these stresses and shocks.[9,10] Common types of environmental products include fuelwood and wild foods.[11] Foraging for environmental products primarily contributes economic value to a household's livelihood portfolio from direct use, rather than from sales.[12] This refers to a product's direct use value (DUV), which quantifies this direct consumption according to local monetary values.[13] In the study region, commonly consumed wild foods include fruits, herbs and insects[13], which households use to supplement their food security and other basic needs[9,14].

Another income stream valued for its DUV in the study region is agriculture, which is broadly divided into crops and livestock.[11] In some African countries, cropping provides the greatest portion of household income[15], roughly equivalent to full-time employment as a labourer[16]. Very little of this value is, however, converted into cash through sales.[3] Case studies from other villages in Bushbuckridge, South Africa, find that most households grow crops for direct use, but may sell some surplus once the household's food security needs are met.[5,16] The money saved by supplementing the household diet with own-grown foods may be invested in other livelihood strategies, such as education.[17]

Households that can afford to buy cattle, or that inherit them, may enjoy agricultural income equivalent to one-third of their entire income portfolio.[17] This income could be generated in regular smaller amounts by selling milk or dung, or by renting cattle out for their draught power. Alternatively, households could 'cash in' their investment by selling the whole animal alive or slaughtered.[17,18] By these means, cattle provide financial stability to a household[18]

EDITOR:
Jane Carruthers

KEYWORDS:
land-based income, off-farm cash income, income streams, rural households, sustainable livelihoods

FUNDING:
National Research Foundation (South Africa); Wellcome Trust (grant 085477/Z/08/Z); Eunice Kennedy Shriver National Institute of Child Health and Human Development

and support other livelihood strategies such as small enterprises, farming and funding an education[17]. Cattle ownership and farming, however, incur considerable expenses such as veterinary bills, hired labour, losses due to theft, and the construction and maintenance of corrals and are therefore associated with wealthier households.[17,18] Households that cannot afford to farm cattle may farm smaller livestock such as goats, which cost less to buy and maintain.[5]

Off-farm incomes, typically from wage labour, non-agricultural self-employment, remittances, social grants, and savings and loans are an important source of cash for rural households.[2,11,19] Globally, off-farm income makes up one quarter (25%) of the income portfolio of rural households. In South Africa, however, off-farm incomes contribute substantially more – up to 47% – to a rural household's income portfolio.[20]

This study sheds new light on the relative importance of different income streams to rural households, drawing from a South African case study site, Bushbuckridge. Existing research deals with the same income streams introduced here,[2,4,9-11,17-19] but none thus far analyse this full income portfolio simultaneously across time, as undertaken here, in order to examine how households optimise the value of their income streams. In this study, we form an understanding of how a household organises and optimises the entire spectrum of value that flows into it from the full range of income sources, be it from foraging, cropping, livestock keeping or cash incomes. The way we do so is to analyse each income stream at three different 'points of analysis' (POA), namely the value flowing in, the portion of this income that is used directly, and the portion that is converted into cash, as described in more detail under the Data Analysis subsection. By analysing all of the household income streams, as well as distinguishing between cash generated and direct use, we demonstrate that, notwithstanding the presence of cropping, foraging and livestock-keeping in the area, off-farm incomes are almost the sole source of cash for households in Bushbuckridge. This key finding not only highlights the extent of deagrarianisation of former homeland regions such as Bushbuckridge, but also challenges what it means to be rural in the South African context.

Methods

Geography of the area

Data collection took place in the Bushbuckridge Local Municipality, Mpumalanga Province, in the northeast corner of South Africa during 2010 (Figure 1). This area is in the savanna biome, characterised by Granite Lowveld bushveld dominated by broad-leafed *Combretaceae* tree species in the sandy uplands, and species such as *Dichrostachys cinerea* and *Grewia bicolor* in the brackish bottomlands.[21] Most of the woody vegetation is between 2 m and 5 m high, with canopy cover varying from 5% in the open lands to 60% in woodlands.[22] The landscape consists of gently undulating hills.[22] The mean annual temperature is 22 °C, with night temperatures rarely low enough to cause frost.[22] The highest levels of rainfall occur in the summer months of October to May, with average annual rainfall varying from 800 mm in the west to 580 mm in the more arid east.[23]

Villages in Bushbuckridge are scattered throughout the landscape, with homesteads aggregated into large settlements surrounded by communal rangelands. A homestead typically features dwellings, animal pens and home gardens in which crops such as maize, squash and ground nuts are cultivated.[8] Rangelands are used for grazing livestock and foraging for environmental products.[9] These practices appear, however, to be undermining local ecological health and the natural resource base.[8]

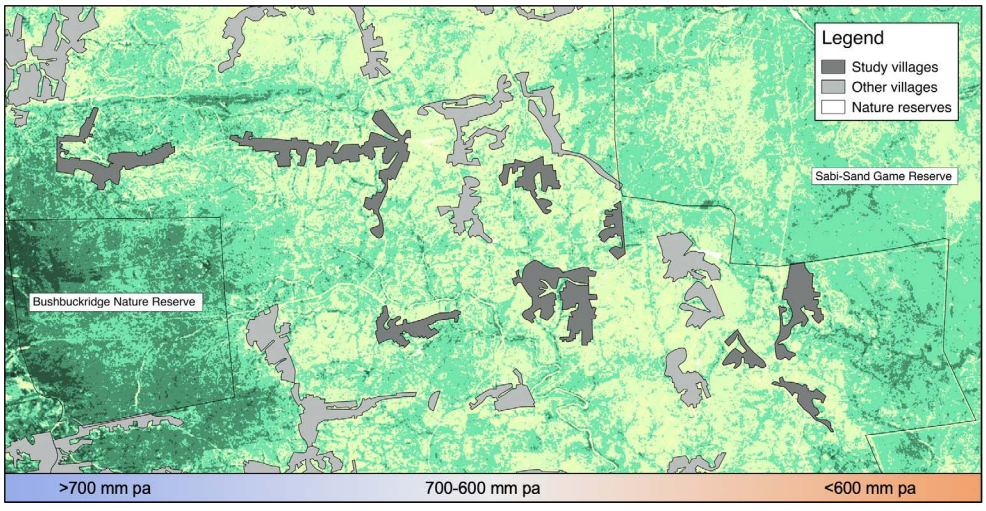


Figure 1: Map of the study site indicating study villages and their environmental context (vegetation cover and position along the rainfall gradient).

Socio-economic characteristics of the population

Under the apartheid government, what is now the Bushbuckridge Local Municipality was divided between two homelands, Gazankulu and Lebowa. The government forcibly moved black South Africans into these homelands in the 1960s and 1970s, resulting in a large increase in human density and thus substantially increasing pressure on natural resources. This has contributed to creating the peri-urban nature of the Bushbuckridge municipality that we see today.[6]

Bushbuckridge has a population of 541 248, according to the latest available census data conducted in 2011, with an average annual growth rate of 0.79% (2001–2011).[24] Population densities remain high, ranging from 146 people/km^2 to over 300 people/km^2.[25] Key challenges to households include low income levels, high food insecurity[25] and an unemployment rate of 50.1%[24]. A large portion of the population (17%) reported earning no income in the latest census, while an additional 24% earned between ZAR1.00 (USD0.14) and ZAR9600.00 (USD1315.07), which is near or below the national poverty line of ZAR7440.00 (USD1019.18) at the time (conversions are based on the mean 2010 exchange rate of USD1.00 = ZAR7.30).[24] Infrastructure in the area, such as roads and piped water, is inadequate, while electricity, although available, is unaffordable to most households for uses other than lighting.[25]

The quality of formal education in Bushbuckridge is low and few tertiary opportunities are available.[25] Although health-care programmes have increased in the area, the growing rate of migrant labour is linked to a high incidence of TB and HIV, which put strain on local health services and household livelihood resilience in general.[25]

The field site, sampling and data collection

Data for this study were collected by teams led by some of the authors for the Sustainability in Communal Socio-Ecological Systems (SUCSES) research project as part of the broader MRC/Wits Agincourt Health and Socio-demographic Surveillance System. The SUCSES research project is a household livelihoods study linked with environmental monitoring, focusing specifically on interactions between humans and their environment over time. Key questions for the interviews dealt with livelihood capital and income-earning activities and correspond with the income streams addressed in this paper. The human ethics clearance for the SUCSES research project was obtained from the Human Research Ethics Committee (Non-Medical) of the University of the Witwatersrand (M10301).

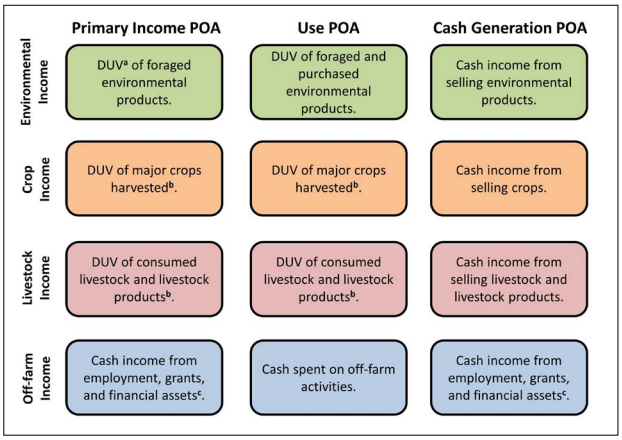
The study area was delineated into three rainfall zones: >700 mm per annum, 600–700 mm per annum and <600 mm per annum, as rainfall directly affects three of the four income streams, namely environmental income, crop income and livestock income. A selection of nine villages was made: three from each rainfall zone to represent the village size range per rainfall zone. Within each rainfall zone, the villages that were the largest, smallest and closest to the mean were selected to ensure that variability in the important contextual factors of village size and rainfall were captured. This population frame included 7502 households. An 8% sample of households was randomly selected from each village in 2010, resulting in a total sample of 590 households. Only two households declined to participate. Interviews took place at the end of the growing season, April–June. Mean household size in our sample ±s.d. was 8.14±4.14.

Face-to-face interviews with an adult representative of each household collected quantitative data using a hardcopy questionnaire that focused on the livelihood capitals, activities, outcomes and incomes of the household. Interviews were conducted in the local language (Shangaan) by experienced local fieldworkers. Each interview lasted about one hour. Informed consent was obtained before every interview and the principles of the Declaration of Helsinki were adhered to.

The data for our study indicated that households drew on four broad livelihood income streams: environmental income, crops, livestock and off-farm incomes. Environmental, crop and livestock incomes include 'in-kind' uses, or the direct use of these products, as well as the cash generated from their sale. Off-farm incomes, however, refer exclusively to cash incomes primarily consisting of migrant labour, remittances, and savings and loans. Agricultural income streams were calculated per annum, due to the seasonal rhythms of income generation characteristic of agriculture, but the other three income streams were recorded as monthly values, multiplied by the months of the year during which such income was obtained to determine annual amounts.

Data analysis

The four income streams (environmental, crops, livestock, off-farm) were analysed using three points that facilitate examination of how households organise and optimise the value of the various streams (Figure 2). The first column in Figure 2, is the *primary income* POA that comprises the DUV of all products and materials entering the household from any source other than cash purchasing. As such, DUV refers to the cash income earned or received from off-farm sources in addition to the cash equivalent of foraged environmental products, crops harvested and livestock products generated. Thus, DUV excludes cash generated by selling foraged environmental products, crops harvested and livestock products. The *use value* POA, the second column of Figure 2, focuses only on that which is consumed by the household. This would include, for example, part (or all) of the goods derived from the primary income, in addition to goods that are purchased for household consumption. Finally, the third column, *cash generation* POA, focuses on the cash derived from selling any left-over products and resources from the primary income (that which the household did not consume directly) as well as cash income earned or received from off-farm sources. As noted, the POAs are not mutually exclusive. The purpose of applying these POAs is not to subdivide the four income streams, but to create a more nuanced understanding of household strategies and the relative contributions of various income streams to overall household portfolios.



[a]*DUV (direct use value): The monetary value of resources used domestically.*[13]

[b]*Households in this study rarely purchased crop or livestock products. This study therefore only includes the DUV of crop and livestock products, and not the purchased value, for evaluating the primary income and use POA.*

[c]*By definition, off-farm primary income POA was the same as off-farm cash generation POA.*

Figure 2: The four income streams (environmental, livestock, crop, off-farm) at three points of analysis (POA).

In areas of high unemployment, the opportunity costs for some household members' time are difficult to calculate, and may be relatively insignificant.[19] Thus, the calculations did not include the opportunity cost of labour spent on foraging or farming. Also excluded was the value of households' utensils and infrastructure, as these did not constitute income flows.

Calculating annual direct use values

Much of the income rural households derive from environmental products, crops and livestock is in the form of direct use, rather than through cash sales.[26] Therefore, the annual DUV of each income stream had to be calculated so that the economic contribution of each income stream to the household could be compared. The DUV for each product was calculated by averaging the different actual prices that households had paid for each resource in the past year.

The local units of measurement were converted to metric. These local measurements included mugs full, bucket loads (5-L, 10-L and 20-L buckets were used), maize meal bags full (50-kg and 80-kg bags were used), bundles, wheelbarrow loads and *bakkie* (pick-up truck) loads. The

buckets were already in litre metrics, and the litre capacity of the maize meal bags is uniform, but we drew on existing research for the metrics for mugs full (0.4 L), *bakkie* loads of fuelwood (532.0 kg), bundles of thatching grass (with a mean diameter of 15 cm) and wheelbarrow loads of fuelwood (39.6 kg).[13,27-30]

To calculate the DUV of environmental products, respondents were asked the volume per month of each resource and for how many months of the year they foraged for it. Residents who had bought environmental products stated the prices they paid per local unit of measurement. The proportion of quotes per number of users for insects (68.4%), fish (58.0%) and fuelwood (19.1%) were adequately large. The proportion of quotes per user for wild fruit (3.5%) and wild vegetables/herbs (0.4%) was very low, but the quoted prices were similar enough to be accepted.

For calculating crop DUV, the same local metrics were applied as with environmental product DUV calculations. Thirty types of crop were identified in the study area, with six identified as 'major' crops for the area, namely: maize (*Zea mays*), peanuts (*Arachis hypogaea*), pumpkin leaf (*Cucurbita pep*), pumpkin (*Cucurbita pep*), bambara beans (*Voandzeia subteranea*) and cowpea (*Vigna unguiculata*). Only the annual volumes of major crops were recorded. Few households sold crops. Thus, even the major crops had no obvious cash value, except maize, which was commercially available and had therefore a relatively stable market value. For this reason, we applied the local estimated price of maize of ZAR1.90 per litre (USD0.26 per litre) to all of the crops, which was inflated from the 1999 prices in Dovie and Shackleton[27] to a 2010 value using a consumer price index of 6% per annum. As the studied households did not report buying crops from one another, the use value and the primary income value were the same for crops (Figure 2). Cattle, goats and pigs were identified as the main types of livestock kept by households in the study area. The DUV per animal for Bushbuckridge is provided by Dovie et al.[6] which we inflated to 2010 values.

Chickens in Bushbuckridge are left to roam free, so their numbers are erratic over the course of a year. These freely roaming chickens were included in the pilot study and, although 58.5% of households reported owning such chickens, few households could quantify the number of chickens that they owned. As it was not possible to quantify the number of chickens a household owned, the chicken products a household consumed could not be categorised as cultivated, bought or foraged. These freely roaming chickens were therefore excluded from the study, although purchased chicken products were included in the study. Excluding the free-roaming chickens may render the findings of household income values from livestock slightly conservative.

The three POAs provide insight into how households use different income streams. This insight allowed us to identify, not only the relative value of the different income streams to household income portfolios, but also to identify which types of income streams are important at which POA. Distinguishing between incomes in this way is important for understanding how households in Bushbuckridge make a living, and the implications this may have for the people, as well as for the local ecology.

Results

Environmental income

Environmental products provided a staple income source for the study's households. The most common such products included wild vegetables/herbs (consumed by 95.7% of the households) and fuelwood (92.2%), as well as wild fruit (53.1%) and edible insects (51.0%).

Almost every household (97.3%) foraged for their own environmental products, accruing a mean annual value of ZAR4383.29 (USD600.45) (Table 1). A few of these households sold environmental products (8.5%), generating a relatively high mean annual income of ZAR12 956.84 (USD1774.91). However, the vast majority of households that sold environmental products made a negligible income (Figure 3), suggesting that only a handful of households were trading environmental products at a commercial scale.

About half of the households (49.2%) bought environmental products, spending ZAR4545.20 (USD622.63) on environmental products per annum. The most common purchases were insects (bought by 35.1% of households) and fuelwood (17.6%). The discrepancy between the value sold and the value bought may indicate that some products were sold or bought outside of the study area.

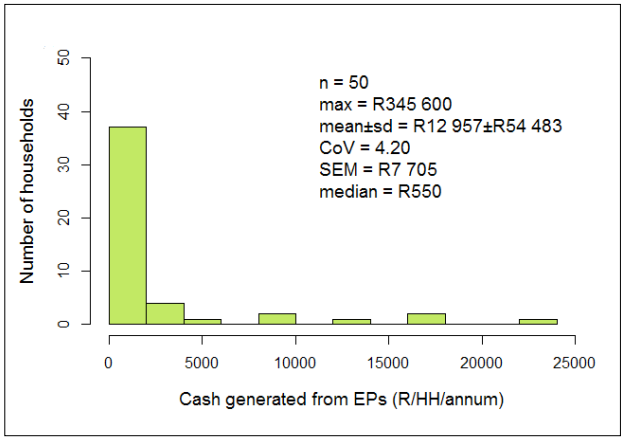


Figure 3: Values of cash (ZAR) generated from selling environmental products (EPs) per household (HH) per annum across a sample size of 590 households in 2010, in Bushbuckridge, South Africa

Table 1: Direct use value† of environmental products at the Use and Cash Generation points of analysis

	Foraged		Bought		Use††		Cash generation	
	Foraging households	All households	Buyer households	All households	User households	All households	Seller households	All households
n	574	590	290	590	580	590	50	590
Mean (ZAR)	4383.29	4264.44	4545.20	2234.09	6683.59	6570.29	12 956.84	1098.066
Standard deviation (ZAR)	6583.29	6532.19	10 055.97	7401.98	9873.69	9827.55	54 482.97	16 124.17
Coefficient of variation	10.95	11.17	16.13	24.16	10.80	10.95	30.66	107.164
Standard error of mean (ZAR)	274.77	268.93	590.50	304.70	409.97	404.57	7705.08	663.79

†Direct use value is the financial value of resources used domestically.[13]

††The number of households using environmental products is not necessarily the sum of households that bought and households that foraged, as the same household may buy and forage, or two different households may share a product that only one of them bought or foraged.

Crop income

Cropping was a popular livelihood activity in Bushbuckridge, with 96.8% practising cropping (Table 2). Households grew a great variety of crops. Thirty types were included in the questionnaire, but only a few types were found in the majority of households' plots. Almost every household (94.6%) grew maize; also popular were peanuts (85.0%), pumpkin leaves (77.4%), pumpkins (66.2%), bambara beans (63.4%), mango (59.0%), cowpea (58.2%) and spinach (52.4%). Spinach was the most popular crop for generating a cash income among the few households (3.7%) that reported selling any of their crops.

Table 2: Direct use value of crops at Use and Cash Generation points of analysis

	Use		Cash generation	
	Cropping households	All households	Seller households	All households
n	549	588	21	588
Mean (ZAR)	615.10	574.29	327.62	11.68
Standard deviation (ZAR)	1776.89	1723.68	273.97	79.13
Coefficient of variation	21.10	21.90	6.13	49.35
Standard error of mean (ZAR)	75.85	71.10	59.79	3.29

Cash sales of crops generated very little actual cash income; the household mean annual cash income among households that sold crops was only ZAR327.62 (USD44.88). Directly using crops was not economically rewarding either: the annual mean value of crops a household consumed directly was ZAR615.10 (USD84.26). This figure is, however, inflated by a minority of households that used relatively greater values of crops; many of the households used less than one-third of this value.

Livestock income

Few households (18.5%) in Bushbuckridge owned livestock. Of those that did, cattle and goats were almost equally popular (owned by 10.8% and 9.5% of households, respectively). The DUV that households gained from consuming their own livestock's products amounted to a mean annual income of ZAR37 444.84 (USD5129.43) (Table 3).

Table 3: Direct use value of livestock products at Use and Cash Generation points of analysis

	Use		Cash generation	
	Owning households	All households	Seller households	All households
n	70	590	47	590
Mean (ZAR)	37 444.84	4442.63	3151.41	251.05
Standard deviation (ZAR)	37 536.45	17661.40	6674.68	2051.52
Coefficient of variation	7.30	29.05	15.48	59.64
Standard error of mean (ZAR)	4486.43	727.08	973.60	84.46

An insignificant number of households bought livestock products, so the households using such products were generally those that owned livestock. Nevertheless, 47 livestock owners (11.9% of all households) reported having sold livestock products in the past year, earning a mean income of ZAR3151.41 (USD431.70) (Table 3). This income is, however, inflated by a minority of high-income earners (Figure 4). The value of livestock sales for the vast majority of livestock owners was negligible, reflected in the median value of ZAR799.35 (USD109.50) for a year's worth of sales, with the majority of sellers earning ZAR199.95 (USD27.39) per annum, at best (Figure 4). This suggests that livestock is only lucrative as an income stream for a small fraction of households in this study.

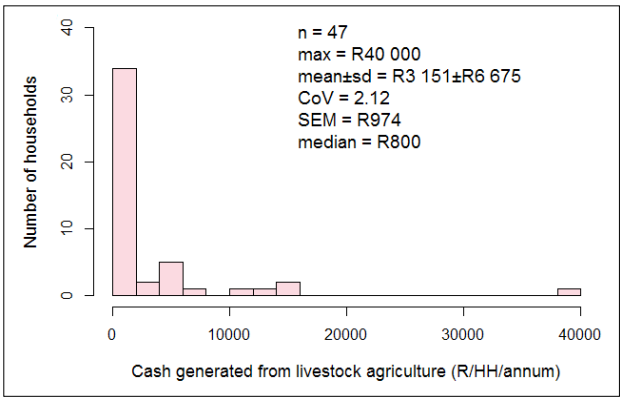


Figure 4: Values of cash (ZAR) generated from selling livestock and livestock products per household (HH) per annum across a sample size of 590 households in 2010, in Bushbuckridge, South Africa

Off-farm income

Off-farm incomes were ubiquitous, with 99.3% of households reporting an off-farm income. The vast majority of households drew on social grants (84.2%), savings and loans (82.0%) and wage employment (local and migrant labour) (73.1%), although it must be noted that the wage employment included temporary employment. Off-farm income was by far the most lucrative of all the income streams studied, with a mean annual household income of ZAR50 514.98 (USD6919.86) (Table 4).

Off-farm income was still relatively low for the majority of the population, with most households falling in the lowest income brackets (Figure 5). Nevertheless, earnings are still far larger than for cash generated by any other income stream. Every household in this study spent cash. The mean annual cash a household spent was ZAR21 178.91 (USD2901.22) – considerably less than the annual mean generated (Figure 5).

Table 4: Direct use value of off-farm cash income at Use and Cash Generation points of analysis

	Use†	Cash generation	
	Spending (all) households	Earning households	All households
n	588	584	588
Mean (ZAR)	21 178.91	50 514.98	50 171.37
Standard deviation (ZAR)	22 046.15	59 278.99	59 222.78
Coefficient of variation	7.59	8.54	8.61
Standard error of mean (ZAR)	909.00	2453.02	2442.29

†'Use' refers to the cash spent on off-farm activities (see Figure 2).

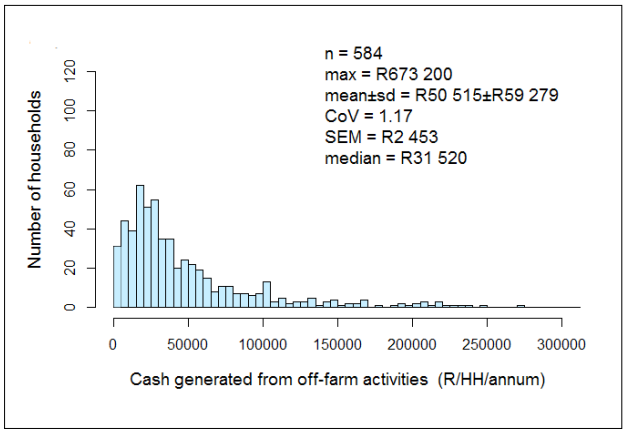


Figure 5: Values of cash (ZAR) generated through wages, social grants and other off-farm activities per household (HH) per annum across a sample size of 590 households in 2010, in Bushbuckridge, South Africa

The complete livelihood portfolio

A central contribution of the present study is examination of livelihood income streams at three POAs to illuminate the relative importance of various types of income. Figure 6 illustrates the relative value that each income stream brings to the study population as a whole (averaged across all households) (n=590), at each POA. As shown by Figure 6, off-farm income contributes the greatest amount of value to the study population.

At the primary income level (Figure 6a), off-farm cash incomes make up 84.4% of the study population's mean annual income. The remainder is evenly distributed between the DUV of environmental products and livestock. While almost every household grows crops, the DUV of crops is relatively insignificant compared to the DUV of environmental products and livestock for primary income.

At the used POA (Figure 6b), crops are once again almost insignificant. The DUV of livestock and environmental products play an important role in total household consumption. Environmental products (both bought and foraged) contribute 20.1% of the value of household income used, while livestock (owned and used) contributes 13.6%. Money spent on off-farm activities makes up the difference, showing the high level of monetisation of Bushbuckridge livelihoods, where over two-thirds of what households consume is purchased.

The cash-income POA (Figure 6c) portrays how important off-farm cash generation is for the study households. Very little cash is generated by selling environmental products, livestock products or crops; off-farm incomes contribute almost the sole source of cash income for the villages in these areas.

Discussion

Land-based income streams have been considered a major part of rural income portfolios.[17,20] However, by dividing the four income streams into different POAs, we were able to identify which land-based incomes played a key role at which POA, as well as to identify the primary sources of income to the area as a whole. A weakness intrinsic to such an approach is the difficulty with which market values may be established, particularly where rural markets are insufficiently developed.[26] In Bushbuckridge, however, considerable work already exists upon which we based our evaluations.[5,27,30] Furthermore, markets in Bushbuckridge are, as we argue in this paper, considerably more developed than expected for South African rural areas.

Our findings reflect a high level of dependence on off-farm income, which suggests a deagrarianisation of Bushbuckridge. Existing research indicates broad trends of deagrarianisation across sub-Saharan Africa due to a number of factors, including resource depletion and aspirations to perceived improvements in quality of life made possible through off-farm employment.[2,4] Recent research shows comparable trends reflected in the study area.[3] While the data presented here are now almost 10 years old, this study still yields useful insights on changing livelihoods in South Africa, and, if anything, it underestimates the current degree to which rural households have become dependent on off-farm livelihood sources. It also provides a valuable baseline with which data from more recent survey rounds (2019 being the latest) can be compared to assess the nature and rates of change.

Environmental income

The DUV of environmental products provides a basic income stream to rural households. Using a meta-analysis of 51 case studies from 17 countries from Africa, Asia and Latin America, Vedeld et al.[31] calculated the mean average DUV of environmental products to be 22% of a rural households' income portfolio (or ZAR6263.40 when adjusted to 2010 values) (USD858.00). Our finding that environmental products contribute 20% of the total value of what households consumed in our study area is remarkably similar. Furthermore, the annual value of environmental products foraged by households in the study area, a mean of ZAR4383.29 (USD600.45), also compares favourably to findings from research in rural areas throughout South Africa that found mean annual income values for environmental products ranging between ZAR4409.20 (USD604.00) and ZAR7497.10 (USD1027.00) (adjusted to 2010 values).[13,28]

Although the vast majority of households in our study used environmental products such as fuelwood, wild vegetables/herbs and insects, few of the households generated sizeable profits from selling environmental products. This finding is consistent with Shackleton et al.'s[10] and Thondhlana and Machapondwa's[12] findings elsewhere in South Africa and suggests that the direct use of environmental products is for fuelwood and basic food security. The same is evident in other studies

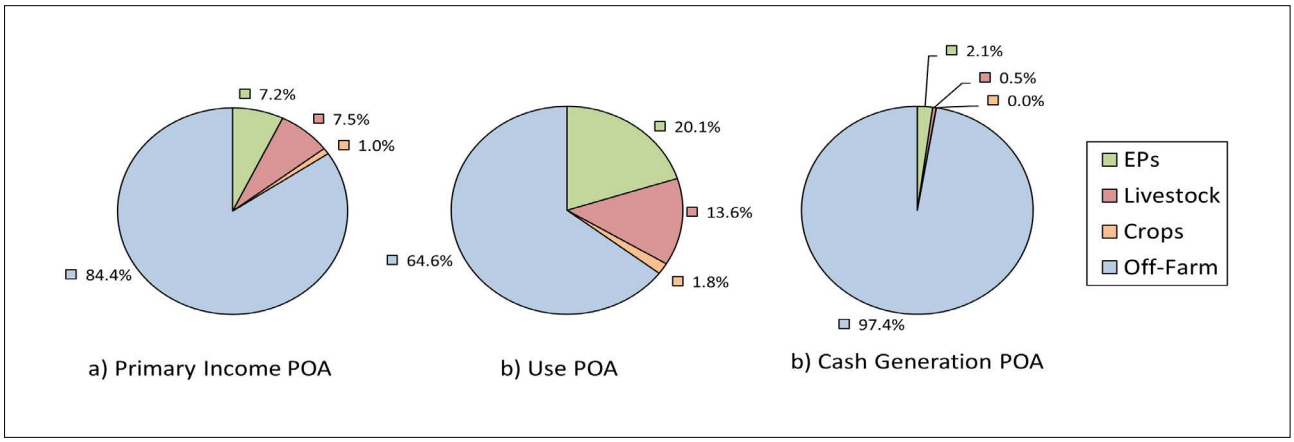


Figure 6: The relative value that each income stream contributes to the mean income portfolio of the study population at the three points of analysis (POA)

from Bushbuckridge.[9,14,26] The few households that were making great profits may be those that use motor vehicles and mechanised equipment to remove large quantities of fuelwood for the market, described by Shackleton et al.[29], Matsika et al.[30] and Cousins[26]. This trend could present a cause for concern if the rate of extraction increases, as the mechanisation of environmental product foraging facilitates the over-exploitation of these natural goods from the commons.[26] Depleted local natural resources may create a poverty trap for those who are marginalised from participation in mainstream economies and who depend largely on the direct use of environmental products for their survival.[1]

Agricultural income

Studies from Africa, Asia and Latin America indicate that cropping and livestock husbandry combined make up 37% of the value of rural households' income portfolios.[31] In the study area however, the combined contribution from these two income streams at the used POA is 15% – less than half of the global average. This discrepancy may be attributable to the low frequency of livestock keeping among households and the small volume or low value of crops grown, as well as rainfall levels below the global average, which may limit agricultural productivity.

Consistent with rural areas elsewhere in Africa,[32] cropping was practised by almost all of the households in the study area. However, unlike other African rural areas in which cropping and environmental income are equal,[15] cropping in Bushbuckridge contributes far less than the other income streams. Even compared to rural villages throughout South Africa,[29] where cropping contributes between 7% and 24% of total household income (ZAR1737.40 to ZAR6964.20 at 2010 values) (respectively, USD238.00 to USD954.00), cropping in our study contributed only 2% of household income when averaged across our study population, and only ZAR615.03 (USD84.25) in DUV to the total value of the income a household uses.

Our finding on the relatively low contribution of cropping to household income portfolios relative to existing research may indicate that reliance on off-farm income has increased in the area since Shackleton et al.'s[29] data were gathered. This finding is supported by more recent research by Shackleton et al.[4] that records a declining trend in cropping as a livelihood throughout South Africa. The low value of cropping in our study may also be due to the constraints on valuating the individual value of diverse crop types: a challenge we overcame by standardising crop value to that of maize. Causality is beyond the scope of this study, but, based on our findings, we would recommend further research on the drivers behind the low cropping incomes in Bushbuckridge.

Research from the area finds that households that are already wealthy can afford to own cattle[17,18], but our findings also show that such households derive considerable value from the DUV of livestock as well (ZAR37 444.84 / USD5129.43). Cash generated by households selling livestock products was more moderate (ZAR3158.27 / USD432.64) than the DUV. Compared with previous research from Bushbuckridge,[29] which found the DUV of livestock contributing ZAR5584.50 (USD765.00) per household per annum, our findings suggest livestock incomes have become lucrative for the few households that keep them.

Off-farm income

Existing research reports that off-farm incomes in rural areas in the developing South in general[31], as well as in South Africa in particular[20], contribute 38% and 47% of household income, respectively. However, our primary income POA reveals that off-farm cash income contributes a far greater proportion (84%) of the value of household income in the study area. Furthermore, its value as a source of cash becomes apparent when compared to the monetary contribution made by the other income streams: off-farm incomes are almost the sole source of cash for households in Bushbuckridge.

Research done in 2000 described a growing trend in cash-based, off-farm transactions by households in Bushbuckridge.[8] Our research confirms this trend, and even suggests that households are far more dependent on off-farm income than indicated by past studies in the area and in the region. A concern is that increasing monetisation could undermine livelihood resilience for the poorest, as a growing cash economy correlates with high resource extraction.[5] Furthermore, research from the study area finds that only the relatively wealthier households can afford for some members to migrate for work.[6] Our findings suggest trends of accrual of money, particularly through employment, to those who are already better educated and wealthier. Coupled with decreasing employment opportunities in South Africa, the off-farm employment opportunities enjoyed by the wealthier households in Bushbuckridge suggest a risk that inequality may become reinforced in Bushbuckridge over time.

Conclusion

The key finding of this paper is the inordinately high reliance on off-farm income among rural households in Bushbuckridge. Cursorily, Bushbuckridge livelihoods appear to be diverse, but by dividing the portfolio into different POAs, we found that off-farm income not only takes up a far larger proportion of total household income value than previously found in local and regional case studies, but also that off-farm incomes are almost the sole source of cash for households in the study area. Livestock, the only other income stream that provides some notable cash income, is the privilege of a select few.

Therefore the majority of households in Bushbuckridge that are less able to participate in the off-farm cash economy, and that subsist through cropping and foraging, may find living conditions becoming increasingly difficult as livelihood sustainability becomes increasingly dependent on cash-based transactions. Rural development, the mandate of which is to reverse historical poverty, must recognise this challenge if it is to have broad benefits and not simply enrich those who are already better off.

Acknowledgements

We thank the research participants and their communities for their time and cooperation. The MRC/Wits Rural Public Health and Health Transitions Research Unit (Agincourt Unit) is acknowledged for invaluable assistance through provision of fieldworkers, field logistics, and data management support. This research was conducted as part of the Sustainability in Communal Socio-Ecological Systems (SUCSES) study which was made possible through funding from the National Research Foundation (South Africa). Indirect support was provided by the Wellcome Trust (grant 085477/Z/08/Z) through its support of the Agincourt Health and Socio-demographic Surveillance System (AHDSS). This research has also benefitted from support provided by the University of Colorado Population Center (Project 2P2CHD066613-06), funded by the Eunice Kennedy Shriver National Institute of Child Health and Human Development. The content is solely the responsibility of the authors and does not necessarily represent the official views of the University of Colorado, CUPC or NIH.

Competing interests

We declare that there are no competing interests.

Authors' contributions

F.H.R. and D.W.O were responsible for the conception and design of the article, reviewing the literature and writing and revising the manuscript. W.T. and B.F.N.E were responsible for the conception and design of the study, revising the article and student supervision. M.C. and L.M.H. were responsible for conception and design of the study and revising the article. C.V. was responsible for revising the article.

References

1. Cumming GS, Buerkert A, Hoffmann EM, Schlecht E, Van Cramon-Taubadel S, Tscharntke T. Implications of agricultural transitions and urbanization for ecosystem services. Nature. 2014;515:50–57. https://doi.org/10.1038/nature13945

2. Bryceson DF. Deagrarianization and rural employment in sub-Saharan Africa: A sectoral perspective. World Dev. 1996;24(1):97–111. https://doi.org/10.1016/0305-750X(95)00119-W

3. Ward CD, Shackleton CM. Natural resource use, incomes, and poverty along the rural-urban continuum of two medium-sized, South African towns. World Dev. 2016;78:80–93. https://doi.org/10.1016/j.worlddev.2015.10.025

4. Shackleton CM, Mograbi PJ, Drimie S, Fay D, Hebinck P, Hoffman MT, et al. Deactivation of field cultivation in communal areas of South Africa: Patterns, drivers and socio-economic and ecological consequences. Land Use Policy. 2019;82:686–699. https://doi.org/10.1016/j.landusepol.2019.01.009

5. Hull E. The social dynamics of labour shortage in South African small-scale agriculture. World Dev. 2014;59:451–460. https://doi.org/10.1016/j.worlddev.2014.02.003

6. Dovie DBK, Witowski ETF, Shackleton CM. Monetary valuation of livelihoods for understanding the composition and complexity of rural households. Agric Hum Values. 2005;22(1):87–103. https://doi.org/10.1007/s10460-004-7233-0

7. Hunter LM, Nawrotzki R, Leyk S, Laurin GJM, Twine W, Collinson M, et al. Rural outmigration, natural capital, and livelihoods in South Africa. Popul Space Place. 2014;20(5):402–420. https://doi.org/10.1002/psp.1776

8. Giannecchini M, Twine W, Vogel C. Land-cover change and human-environment interactions in a rural cultural landscape in South Africa. Geogr J. 2007;173(1):26–42. Available from: https://www.jstor.org/stable/30113491

9. Hunter LM, Twine W, Patterson L. "Locusts are now our beef": Adult mortality and household dietary use of local environmental resources in rural South Africa. Scand J Public Health. 2007;69:165–174. https://doi.org/10.1080/14034950701356385

10. Shackleton SE, Shackleton CM, Netshiluvhi TR, Geach BS, Balance A, Fairbanks DHK. Use patterns and value of savanna resources in three rural villages in South Africa. Econ Bot. 2002;56(2):130–146. https://doi.org/10.1663/0013-0001(2002)056[0130:UPAVOS]2.0.CO;2

11. Angelsen A, Jagger P, Babigumira R, Belcher B, Hogarth NJ, Bauch S, et al. Environmental income and rural livelihoods: A global-comparative analysis. World Dev. 2014;64(suppl 1):S12–S28. https://doi.org/10.1016/j.worlddev.2014.03.006

12. Thondhlana G, Machapondwa E. Dependence on environmental resources and implications for household welfare: Evidence from the Kalahari drylands, South Africa. Ecol Econ. 2014;108:59–67. https://doi.org/10.1016/j.ecolecon.2014.10.003

13. Twine W, Moshe D, Netshiluvhi T, Siphugu V. Consumption and direct-use values of savanna bio-resources used by rural households in Mametja, a semi-arid area of Limpopo Province, South Africa. S Afr J Sci. 2003;99(9–10):467–473. https://hdl.handle.net/10520/EJC97689

14. Shackleton CM, Shackleton SE. The importance of non-timber forest products in rural livelihood security and as safety nets: A review of evidence from South Africa. S Afr J Sci. 2004;100:658–664. https://hdl.handle.net/10520/EJC96169

15. Pouliot M, Treue T, Obiri BD, Ouedraogo B. Deforestation and the limited contribution of forests to rural livelihoods in West Africa: Evidence from Burkina Faso and Ghana. Ambio. 2012;41(11–12):738–750. https://doi.org/10.1007/s13280-012-0292-3

16. High C, Shackleton CM. The comparative value of wild and domestic plants in home gardens of a South African rural village. Agrofor Syst. 2000;48(2):141–156. https://doi.org/10.1023/A:1006247614579

17. Chaminuka P, Udo HMJ, Eilers KCHAM, Van der Zijpp A. Livelihood roles of cattle and prospects for alternative land uses at the wildlife/livestock interface in South Africa. Land Use Policy. 2014;38:80–90. https://doi.org/10.1016/j.landusepol.2013.10.007

18. Shackleton CM, Shackleton SE, Netshiluvhi TR, Mathabela FR. The contribution and direct-use value of livestock to rural livelihoods in the Sand River catchment, South Africa. Afr J Range For Sci. 2005;22:127–140. https://doi.org/10.2989/10220110509485870

19. Babulo B, Muys B, Nega F, Tollens E, Nyssen J, Deckers J, et al. Household livelihood strategies and forest dependence in the highlands of Tigray, Northern Ethiopia. Agric Syst. 2008;98(2):147–155. https://doi.org/10.1016/j.agsy.2008.06.001

20. Mugido W, Shackleton CM. The contribution of NTFP trade to rural livelihoods in different agro-ecological zones of South Africa. Int For Rev. 2017;19(3):306–320. https://doi.org/10.1505/146554817821865063

21. Mucina L, Rutherford MC, editors. The vegetation of South Africa, Lesotho and Swaziland. Strelitzia 19. Pretoria: South African National Biodiversity Institute; 2006.

22. Venter FJ, Scholes RJ, Eckhardt HC. The abiotic template and its associated vegetation pattern. In: Du Toit KH, Biggs J, Rogers H, editors. The Kruger experience: Ecology and management of savanna heterogeneity. London: Island Press; 2003. p. 83–129.

23. Shackleton CM. Comparison of plant diversity in protected and communal lands in the Bushbuckridge lowveld savanna, South Africa. Biol Conserv. 2000;94:273–285. http://hdl.handle.net/10204/1497

24. StatsSA. Bushbuckridge [webpage on the Internet]. c2011 [cited 2019 Aug 28]. Available from: http://www.statssa.gov.za/?page_id=993&id=bushbuckridge-municipality

25. Kahn K, Collinson MA, Gomez-Olive FX, Mokoena O, Twine R, Mee P, et al. Profile: Agincourt Health and Socio-demographic Surveillance System. Int J Epidemiol. 2012;41(4):988–1001. https://doi.org/10.1093/ije/dys115

26. Cousins B. Invisible capital: The contribution of communal rangelands to rural livelihoods in South Africa. Dev South Afr. 1999;16(2):299–318. http://dx.doi.org/10.1080/03768359908440079

27. Dovie DBK, Shackleton CM. Direct-use value of smallholder crop production in a semi-arid rural South African village. Agric Syst. 2003;76(1):337–357. https://doi.org/10.1016/S0308-521X(02)00124-5

28. Dovie DBK, Shackleton CM, Witkowski TF. Direct-use values of woodland resources consumed and traded in a South African village. Int J Sus Dev World. 2002;9(3):269–283. https://doi.org/10.1080/13504500209470122

29. Shackleton SE, Kaschula S, Twine WC, Hunter L, Holding-Anyonge C, Petheram L. Forests as safety nets for mitigating the impacts of HIV/AIDS in southern Africa. Forest Livelihood Briefs. 2006;4:1–4 [cited 2018 Apr 26]. Available from: http://www.cifor.org/publications/pdf_files/livebrief/livebrief0604e.pdf

30. Matsika R, Erasmus BFN, Twine WC. Double jeopardy: The dichotomy of fuelwood use in rural South Africa. Energy Policy. 2013;52:716–725. https://doi.org/10.1016/j.enpol.2012.10.030

31. Vedeld P, Arild A, Bojo J, Sjaastad E, Berg GK. Forest environmental incomes and the rural poor. For Pol Econ. 2007;9(7):869–879. https://doi.org/10.1016/j.forpol.2006.05.008

32. Babulo B, Muys B, Nega F, Tollens E, Nyssen J, Deckers J, et al. The economic contribution of forest resource use to rural livelihoods in Tigray, Northern Ethiopia. For Pol Econ. 2009;11(2):109–117. https://doi.org/10.1016/j.forpol.2008.10.007

AUTHORS:
Jessica Leaver[1]
Michael I. Cherry[1]

AFFILIATION:
[1]Department of Botany and Zoology, Stellenbosch University, Stellenbosch, South Africa

CORRESPONDENCE TO:
Jessica Leaver

EMAIL:
jes.leaver@gmail.com

DATES:
Received: 22 Oct. 2019
Revised: 18 Mar. 2020
Accepted: 02 Apr. 2020
Published: 29 Sep. 2020

HOW TO CITE:
Leaver J, Cherry MI. Forest product harvesting in the Eastern Cape, South Africa: Impacts on habitat structure. S Afr J Sci. 2020;116(9/10), Art. #7508, 9 pages. https://doi.org/10.17159/sajs.2020/7508

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:
medical bark harvesting, timber harvesting, pole harvesting, habitat degradation, habitat modification

FUNDING:
South Africa National Research Foundation through the Foundational Biodiversity Information Programme (FBIP 98871)

Forest product harvesting in the Eastern Cape, South Africa: Impacts on habitat structure

The Eastern Cape Province harbours 46% of South Africa's remaining indigenous forest cover, and is one of the country's poorest and least developed provinces. Forest resources thus represent a vital component of rural livelihoods in this region. Consequently, forest management policies aim to balance the needs of resource users with the ecological integrity of forest ecosystems. In a recent study, forest bird ranges were shown to have declined in the Eastern Cape over the past 20 years, despite increases in forest cover over the same time period, indicating that habitat degradation may be driving forest bird losses. Given that harvesting of forest products represents the primary human disturbance in forests in the Eastern Cape today, insight is needed regarding the link between resource use and habitat modification. We report on effects of harvesting of three key forest products – poles, timber and medicinal bark – on habitat structure at the ground, understorey and canopy layers in indigenous forests in the province. Harvest activities had considerable impacts on habitat structure, depending on the nature and extent of harvesting. Bark and timber harvesting resulted in canopy gaps, whereas pole harvesting reduced tree density, resulting in understorey gaps. Overall, harvest activities increased the frequency of canopy disturbance, and density of understorey layer foliage. Unsustainable bark harvesting practices increased the mortality rate of canopy trees, thereby increasing dead wood availability. By providing insight into human-mediated habitat modification in forests of the Eastern Cape, this study contributes to the development of ecologically informed sustainable resource management policies.

Significance:

- Unregulated harvesting of forest products in state-managed indigenous forests of the Eastern Cape results in habitat modification.
- The nature and extent of habitat modification is dependent on the type and intensity of resource use, indicating that resource use may be sustainably managed.
- Timber and medicinal bark harvesting activities result in canopy disturbances, thereby altering natural canopy gap dynamics, with concomitant impacts on understorey habitat structure.
- Changes in forest habitat structure associated with high levels of resource use are likely to have ramifying effects on forest biodiversity.

Introduction

Habitat loss and modification are currently the primary drivers of forest biodiversity loss globally.[1] Unlike many parts of Africa, forest cover in the Eastern Cape, which harbours close to half (46%) of South Africa's remaining indigenous forest cover, has increased over the past 20 years[2] – an increase which is attributed to the revegetation of previously cultivated fields in response to increasing trends of de-agrarianisation in rural areas[3], together with carbon fertilisation[4]. Thus, while habitat loss appears not to be a major threat to forest biodiversity, degradation has been identified as a major concern.[5-9] While much forest degradation in South Africa is attributed to extensive historical logging[10], commercial-scale logging has not occurred in indigenous forests in the Eastern Cape for the past 80 years, after being outlawed in 1939 in all but one forest complex, where limited commercial harvesting was re-introduced in 1975[9]. Consequently, informal harvesting of forest products now comprises the primary anthropogenic disturbance in forest habitats in the region[5-9] and is largely related to poor socio-economic conditions in the province[11]. Thus, although forests comprise a mere 2.2% of provincial land cover[12], their socio-economic value is significant, with thousands of rural households dependent on forest resources for subsistence and commercial use[13]. While forest policies in South Africa aim to develop forests for sustainable use, several studies have reported unsustainable harvest rates occurring across the region[7,8,12,14], largely attributed to a decline in the capacity of institutional and traditional structures to regulate resource use[15,16]. A *de facto* open-access system thus governs forest resource use in South Africa today, leading to increasing concern that unregulated resource use is degrading forest habitats and compromising the conservation of forest biodiversity.

Long-term harvesting of forest products has significant effects on temperate forest habitats, driving changes in habitat structure and tree species composition, even when occurring at relatively low levels.[8,17] Moreover, the ecological impact of resource use depends on the plant part harvested and intensity of use.[18] Thus, while grazing of livestock in forests may affect soil quality[19] and increase exotic cover[20], timber harvesting affects canopy closure, mean tree size and understorey density[21,22]. The extent to which a resource has been commercialised is also of consequence, as resources used to generate income, particularly in the context of *de facto* open-access systems, are often harvested more intensely, and frequently unsustainably, and thereby have more profound ecological impacts.[12,14,23]

Human activities that modify habitat structure, in turn, may influence faunal community assemblage in forests. For example, habitat features at the local scale relate to the occurrence of specific functional traits and community

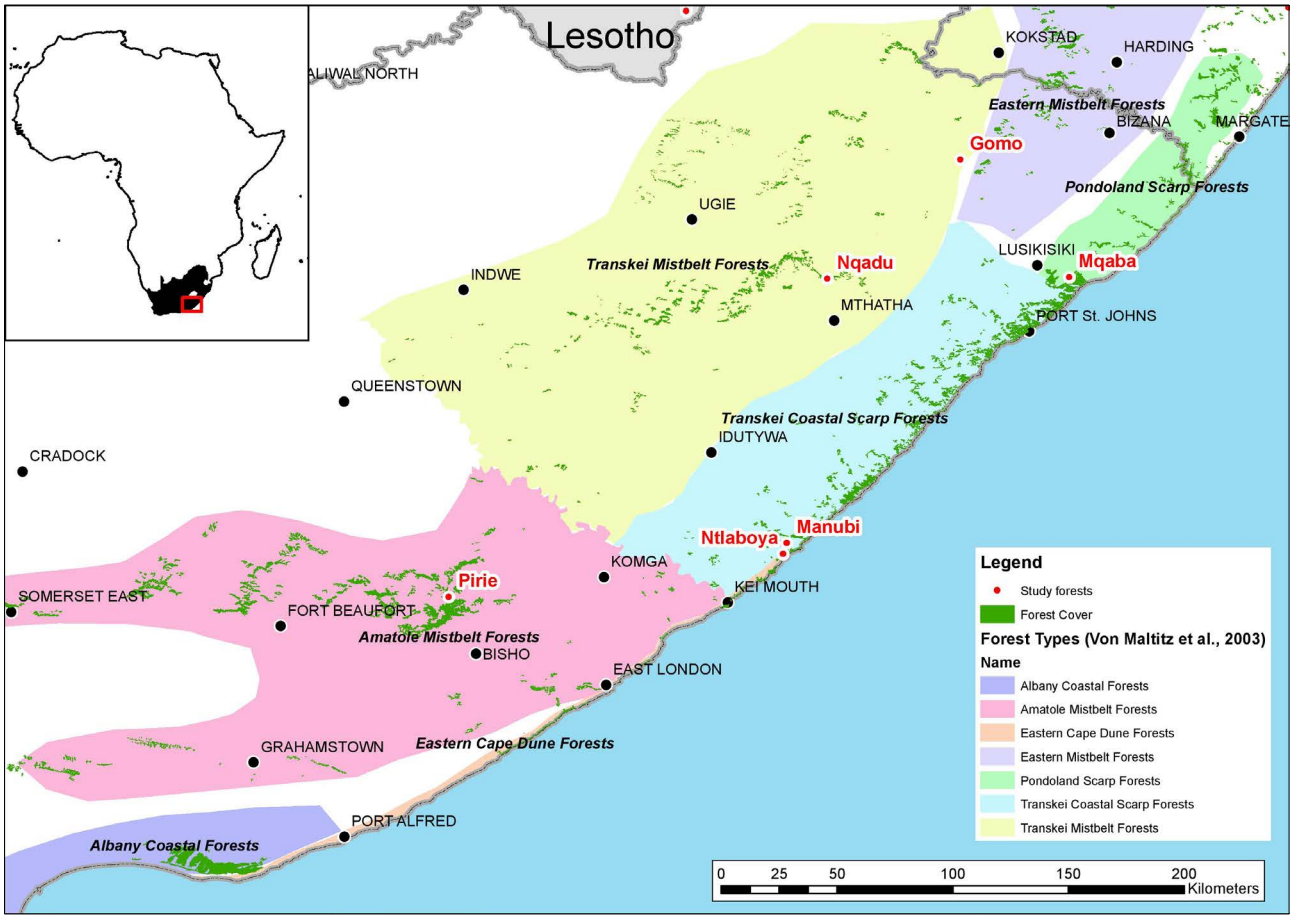
structure in avifaunal populations.[24] Consequently, studies have shown forest faunal populations, including amphibians, bats, birds and reptiles, to be sensitive to human-mediated changes in habitat structure, with species specialised in their foraging or microhabitat requirement being particularly sensitive.[25-30] Given the critical ecosystem functions provided by forest fauna – including seed dispersal, pest control and pollination[31] – human activities that modify habitat structure may have ramifying effects on forest ecosystem functioning.

In a recent study, half of South Africa's forest-dependent bird species were shown to have experienced range declines in the past 20 years, with declines most notable in the Eastern Cape, despite forest cover increases in this region over the same time period.[2] This finding suggests that habitat-scale disturbances rather than landscape-scale habitat loss may be driving bird declines in the region. We thus aimed to assess the effects of harvest activities on habitat structure (defined as the composition and arrangement of physical matter at a location[32]) at the forest scale. Specifically, we examined how different harvest activities modify habitat structure at the canopy, understorey and ground level in six forests, representative of five national forest types, across the Eastern Cape region. Resource use focused on extraction of live biomass from forests, namely understorey trees for poles, canopy trees for timber and crafts, and bark for medicinal purposes, as these represent key resource use types in the region.[8]

Methods

Study site

The study was conducted in the Eastern Cape Province of South Africa between April and July 2016. Forest cover in this region is discontinuous and highly fragmented (Figure 1).[9] Within the study area, six forests were sampled, including five national forest types in the two main zones of forest, i.e. the lowland coastal and scarp forests of the subtropical coastal zone, and the warm-temperate mistbelt forests found on the south to southeastern aspect of inland mountain ranges (Figure 1).[9] Specifically, the following forests were sampled: Mqaba (Pondoland Scarp Forest), Manubi (Transkei Coastal Forest) and Ntlaboya (Eastern Cape Dune Forest) of the lowland zone; and Gomo, Nqadu (Transkei Mistbelt Forest) and Pirie (Amathole Mistbelt Forest) of the montane zone (Figure 1). Within the Transkei mistbelt region, forests located within matrixes of timber plantations leased by the state to private companies are often deemed to be better protected than those which are not, so a forest in each category was sampled, with Nqadu associated with privately managed plantations while Gomo was associated with plantations managed by the South African Department of Environment, Forestry and Fisheries (DEFF). Study forests were selected based on their size, protected status, and the proximity of surrounding human settlements. Specifically, selected forests were greater than 150 ha, and unfenced; managed by DEFF; and had rural settlements within 4 km of the forest boundary. While most forest patches in the Eastern Cape are smaller than 150 ha, and prone to negative effects of fragmentation,[5,9] study forests were selected to represent larger, more 'intact' forest patches within the region. This is because these forests have larger core areas (i.e. portion of forests unaffected by edge effects) and are thus of high biodiversity value, such that insight into anthropogenic pressures within these forests is of conservation priority.[5] Furthermore, given that 70% of forests in the Eastern Cape region are managed by DEFF, and are often associated with communities in close proximity,[5] study forests are representative of the current socio-political context within which larger, 'intact' forests in the region occur. Lastly, study forests have endured colonial logging,[10] followed by subsistence harvesting in recent times, such that they are representative of the history of human impacts.



Source: Forest types after Von Maltitz et al.[9]

Figure 1: Location of the six study forests in the Eastern Cape Province, South Africa.

Study design

A total of 89 circular plots of 0.04 ha (radius 11.3 m) were sampled, with an average of 15 plots sampled per forest. Points for sampling plots in each forest were selected to represent varying levels of disturbance from resource use, based on detailed discussion and guided walks in each forest with DEFF staff (forest managers and/or forest guards), and local community members, in addition to visual assessment by J.L. of human use in each forest, conducted over two reconnaissance trips prior to sampling. Plot locations were selected to represent the continuum of harvesting disturbances present at each forest, from heavily harvested sites to those with little or no harvesting present. Heavily harvested plots were defined as those where >20% of available stems were harvested for poles, timber or bark. Where 10–20% of available stems were harvested, plots were described as intermediately harvested, while low levels of harvesting were defined as harvest levels of <5% of any resource at the plot level. This non-random sampling approach aimed to provide an objective overview of resource use within each forest, as well as samples from the full range of harvest activities and intensities, against which to investigate habitat changes and, in a linked study, avifaunal responses to resource use.[24] Based on this categorisation, 6% (n=8) of plots overall had no harvesting, while 27%, 30% and 34% had low, intermediate and high levels of harvesting, respectively. A minimum distance of 150 m was maintained between selected plots, and 50 m between plots and the forest edge (i.e. all survey sites were within the forest interior), while distance into the forest interior ranged from 50 m to 900 m.

Data collection

At each plot, microhabitat structure and foliage profile were recorded within three nested circular plots of 0.2 ha (radius of 25.2 m), 0.04 ha (radius of 11.3 m), and 0.01 ha (radius of 5.6 m), respectively. In the 0.2 ha plot, all standing dead trees (henceforth, snags) were recorded by diameter (cm) at 1.3 m above the ground, i.e. diameter at breast height (DBH), and cause of death, i.e. natural or due to bark harvesting. Natural snags include standing trees that have died due to factors other than harvesting, such as wind effects, senescence or disease. In the 0.04-ha plot, the following variables were recorded: DBH of all living stems (>5 cm DBH); percentage canopy cover; mean canopy height; percentage coverage of bare ground, leaf litter, grass cover and herbaceous cover; and foliage density at 0–0.5 m; 0.5–1 m; 1–2 m; 2–5 m; 5–10 m and 10–20 m. Foliage density at each height class was estimated using a 8-m-long telescoping pole and marked at each height interval. The pole was sequentially set up at eight evenly spaced points 11.3 m from the plot centre (i.e. along the 0.04-ha circular plot boundary) and visual estimates of foliage density (as a percentage) at each height class were made from the plot centre. Foliage density scores were further converted into a foliage height diversity index (FHDI) using the Shannon–Weiner Diversity Index formula, as follows:

$$H = -\sum_{i=1}^{s} p_i \ln(p_i)$$

where p_i is the proportion of the total foliage which lies in the ith layer of the chosen horizontal layers. This index thus provided a measure of the vertical heterogeneity at each plot.

A rangefinder was used to assist with estimates of foliage density beyond the length of the telescoping pole, as well as to estimate mean canopy height at each plot. Abundance of coarse woody debris was measured based on the number of grounded dead logs (diameter >10 cm; length >1.5 m). Harvest activities were also measured in the 0.04-ha plot: stumps, i.e. trees harvested for poles or timber, were counted and diameter measured. Based on diameter, stumps were categorised as pole (5–19.9 cm diameter) or timber (>20 cm diameter) harvesting, after Obiri et al.[8] Trees harvested for medicinal bark were recorded using DBH and extent of bark removal on individual trees up to 3 m on the tree stem (scored 1 – 6 based on percentage of bark removed, where 1 = 1–10%; 2 = 11–25%; 3 = 26–50%; 4 = 51–75%; 5 = ringbarked to any extent %; 6 = total ringbark, where ringbarked stems are those where bark has been removed from around the full circumference of the stem, after Cunningham[33]. In 0.01-ha plots, sapling (stem diameter 1–5 cm) abundance was recorded.

Data analyses

Pole and timber harvest intensities were calculated per plot for each size class based on the accumulated harvestable stems (stumps plus standing stems) as follows:

$$\text{Tree harvest index}_j = \text{number stumps}_j / (\text{number stumps}_j + \text{number stems}_j),$$

where j represents the size class being assessed.

Bark harvest intensity was assessed based on a bark harvest index derived from summed bark removal scores assigned to individual bark-harvested trees, calculated at each plot, as follows:

$$\text{Bark harvest index} = \text{summed bark removal score/no. individuals bark harvested}$$

Harvest effects on forest structure were investigated using (1) linear mixed models for habitat variables measured on a continuous scale; (2) generalised linear mixed models for habitat variables measured as counts, and (3) beta regression for habitat variables measured as per cent cover. A mixed-modelling approach was used in all cases to account for the nested study design, with sample forests included as a random effect throughout the analysis to account for plots being nested within study forests. Separate models were used to assess the response of each habitat feature to harvesting, with pole, timber and bark harvest indices included as the explanatory variables in addition to, and in all possible combinations of two-way interactions with one another. The two-way interaction between timber and bark was not included, as bark and timber harvesting were seldom recorded within a single plot. Spearman's rank correlation test was used to test for significant correlations between harvest variables, to avoid issues related to multicollinearity. The test showed the harvesting variables to be uncorrelated ($-0.4 < r > 0.4$). Habitat variables measured as counts (tree, snag, sapling and grounded log abundance) were modelled using generalised linear mixed models, with a Poisson distribution and log-link. Response variables measured as per cent cover were converted to proportions and modelled using a beta regression. Model assumptions were verified by plotting residuals versus fitted values, and versus each covariate in the model. Where interaction terms did not improve model strength based on Akaike information criterion (AIC) values, they were removed from the final model. Data from Pirie were not included in these analyses as minimal harvesting was recorded in this forest, and these analyses aimed to assess effects in disturbed forests.

Results

Of the 18 measured structural variables, 12 were significantly impacted by harvesting activities, with responses dependent on the type and intensity of resource use (Figures 2–5; Table 1). Furthermore, the two-way interaction between pole and timber harvesting was shown to affect structural habitat heterogeneity (Figure 5; Table 1). Five habitat features were unaffected by harvest activities: canopy height; mean DBH; lower-understorey (0.5–1 m) foliage density; mid-storey (2–5 m) foliage density; and canopy layer (5–10 m) foliage density (Table 1).

Table 1: Response of structural habitat variables to pole harvesting intensity (Pole), timber harvesting intensity (Timber) and bark harvesting intensity (Bark) derived from linear mixed models; *p*-values in bold indicate statistical significance ($p<0.05$)

Response	Fixed effect	Estimate	s.e.	*t*-value	*p*-value
Foliage height diversity index	Intercept	1.87	0.04	45.87	0.00
	Pole	0.24	0.13	1.85	0.07
	Timber	0.47	0.24	1.97	0.05
	Bark	-0.01	0.04	-0.32	0.75
	Pole*Timber	-3.20	1.41	-2.26	**0.03**
Canopy height (m)	Intercept	12.56	0.91	13.87	0.00
	Pole	-0.14	1.85	-0.08	0.93
	Timber	0.25	3.27	0.08	0.94
	Bark	0.27	0.79	0.33	0.74
Mean diameter at breast height (DBH) (log-transformed)	Intercept	1.25	0.04	31.13	0.00
	Pole	0.06	0.10	0.64	0.52
	Timber	-0.20	0.18	-1.13	0.26
	Bark	-0.02	0.04	-0.51	0.61
Mean basal area per ha (log)	Intercept	1.79	0.06	28.99	0.00
	Pole	-0.42	0.13	-3.22	**<0.01**
	Timber	-0.02	0.23	-0.09	0.93
	Bark	-0.02	0.06	-0.43	0.67
Coarse woody debris	Intercept	1.78	1.17	10.68	0.00
	Pole	-0.82	0.43	-1.88	0.06
	Timber	1.42	0.73	1.94	**0.05**
	Bark	0.50	0.17	3.02	**<0.01**
Tree abundance (>5 cm DBH)	Intercept	4.16	0.11	39.32	0.00
	Pole	-0.91	0.16	-5.83	**<0.01**
	Timber	0.18	0.26	0.68	0.50
	Bark	-0.03	0.07	-0.43	0.68
Sapling abundance (<5 cm DBH)	Intercept	3.63	0.20	18.46	0.00
	Pole	-1.29	0.19	-6.73	**<0.01**
	Timber	0.81	0.27	3.06	**<0.01**
	Bark	-0.34	0.07	-4.49	**<0.01**
Snag abundance (DBH > 10 cm)	Intercept	2.16	0.17	12.93	0.00
	Pole	-0.65	0.38	-1.70	0.09
	Timber	-0.62	0.65	-0.95	0.34
	Bark	0.08	0.14	5.87	**<0.01**
Canopy cover (%)	Intercept	1.06	0.06	17.55	0.00
	Pole	-0.64	0.36	-1.81	0.07
	Timber	-1.41	0.57	-2.45	**<0.05**
	Bark	-0.72	0.14	-5.10	**<0.01**
Herb cover (%)	Intercept	-1.41	0.46	-3.05	0.00
	Pole	2.49	0.81	3.08	**<0.01**
	Timber	-2.22	1.72	-1.30	0.20
	Bark	0.13	0.36	0.35	0.73
Leaf litter cover (%)	Intercept	-0.52	0.30	-1.75	0.00
	Pole	-2.95	0.77	-3.79	**<0.01**
	Timber	-1.45	1.13	-1.28	0.20
	Bark	-0.46	0.30	-1.55	0.12
Bare ground cover (%)	Intercept	-3.17	0.27	-11.87	0.00
	Pole	0.22	0.86	0.26	0.80
	Timber	3.43	1.21	2.84	**<0.01**
	Bark	-0.44	0.33	-1.32	0.19
Herb layer foliage density (0–0.5 m)	Intercept	-0.003	0.48	-0.007	0.99
	Pole	1.72	0.85	2.02	**<0.05**
	Timber	0.36	1.27	0.28	0.77
	Bark	1.48	0.35	4.26	**<0.01**
Lower-understorey foliage density (0.5–1 m)	Intercept	-0.61	0.22	-2.76	0.00
	Pole	0.99	0.78	1.27	0.21
	Timber	1.73	1.37	1.26	0.21
	Bark	0.36	0.32	1.14	0.26
Upper-understorey foliage density (1–2 m)	Intercept	-1.10	0.23	-4.76	0.00
	Pole	0.81	0.82	0.99	0.32
	Timber	3.41	1.37	2.48	**<0.05**
	Bark	-0.20	0.34	-0.59	0.56
Understorey foliage density (0–2 m)	Intercept	-0.53	0.14	-3.68	0.00
	Pole	0.96	0.54	1.76	0.08
	Timber	1.65	0.94	1.75	0.08
	Bark	0.43	0.23	1.93	**0.05**
Mid-storey foliage density (2–5 m)	Intercept	-0.90	0.22	-4.17	0.00
	Pole	-0.22	0.65	-0.34	0.74
	Timber	0.10	1.12	0.09	0.93
	Bark	0.001	0.30	0.00	1.00
Canopy layer foliage density (5–10 m)	Intercept	-0.12	0.18	-0.66	0.51
	Pole	0.78	0.47	1.66	0.10
	Timber	-0.92	0.84	-1.10	0.27
	Bark	-0.24	0.20	-1.19	0.23

Bark harvesting

Increasing bark harvesting intensity negatively affected canopy cover and sapling abundance (<5 cm DBH), while herb layer (0–0.5 m) foliage density, overall understorey (0–2 m) foliage density, number of grounded logs and snag abundance (i.e. standing dead trees; >10 cm DBH) increased with bark harvesting intensity (Figure 2; Table 1).

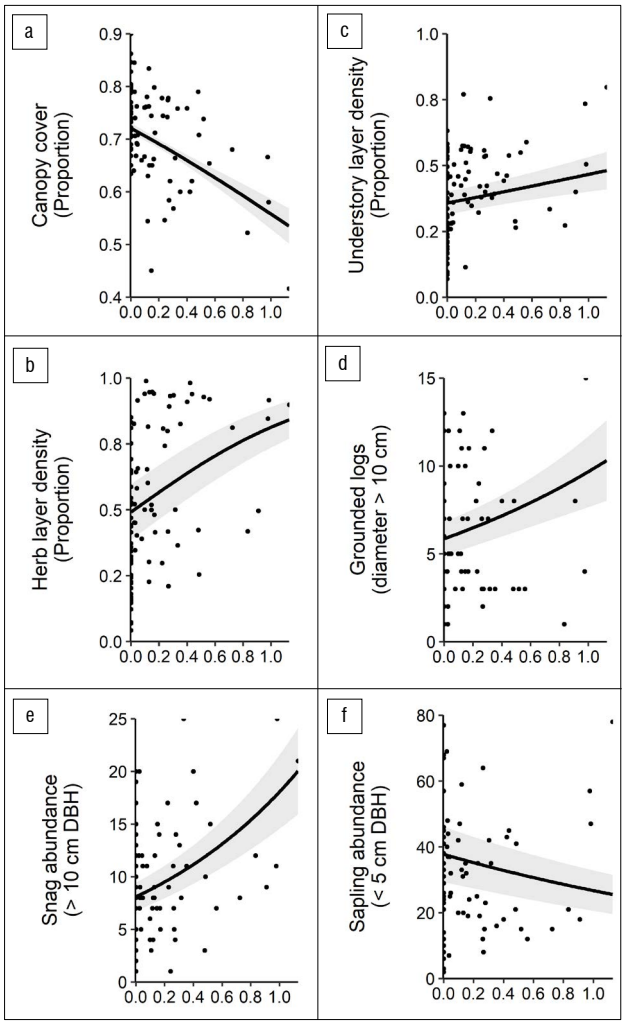


Figure 2: Effect of bark harvesting intensity, measured as the summed score of bark removal from individual bark-harvested trees per plot, based on Cunningham[33] on (a) canopy cover, (b) herb layer (0–0.5 m) foliage density, (c) understorey (0–2 m) foliage density, (d) number of ground logs, (e) snag abundance and (f) sapling abundance (<5 cm diameter at breast height). Relationships shown are derived from mixed models with forest included as a random effect (Table 1); however, graphic representations depict population-level predictions (i.e. excluding random effects).

Pole harvesting

Increasing pole harvesting intensity resulted in declines in tree abundance, sapling abundance, basal area per hectare and leaf litter cover. Conversely, herb layer (0–0.5 m) foliage density and herb cover increased with increasing pole harvesting intensity (Figure 3; Table 1).

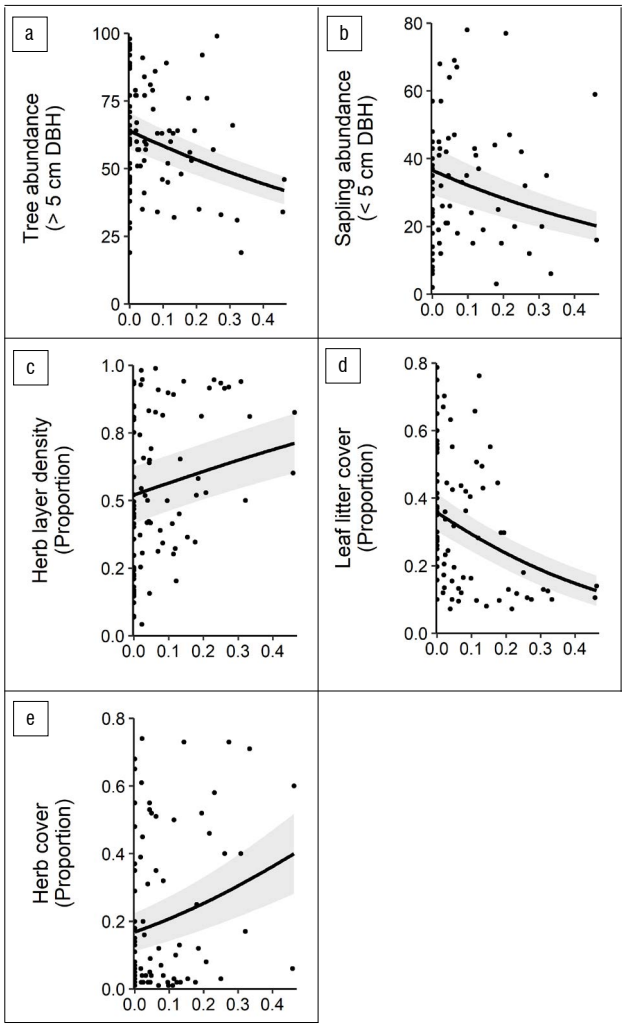


Figure 3: Effect of pole harvesting intensity, measured as the proportion of available stems (5–19.9 cm diameter at breast height/DBH) harvested, on (a) tree abundance (>5 cm DBH), (b) sapling abundance (<5 cm DBH), (c) herb layer (0–0.5 m) foliage density, (d) leaf litter cover and (e) herb cover. Relationships shown are derived from mixed models with forest included as a random effect (Table 1); however, graphic representations depict population-level predictions (i.e. excluding random effects).

Timber harvesting

Increasing timber harvesting intensity resulted in a decline in canopy cover, while a positive relationship was found between the extent of timber harvesting and upper-understorey layer (1–2 m) foliage density, number of grounded logs, bare ground cover, and sapling abundance (Figure 4; Table 1).

Interacting harvest effects

Foliage height diversity index (FHDI) was negatively affected by the interaction between timber and pole harvesting. Specifically, FHDI increased in response to increasing timber harvest intensities where pole harvest levels were low (i.e. 5% of available stems), but declined in response to increasing timber harvest levels where pole harvest intensities were high (i.e. 20% of available stems; Figure 5; Table 1).

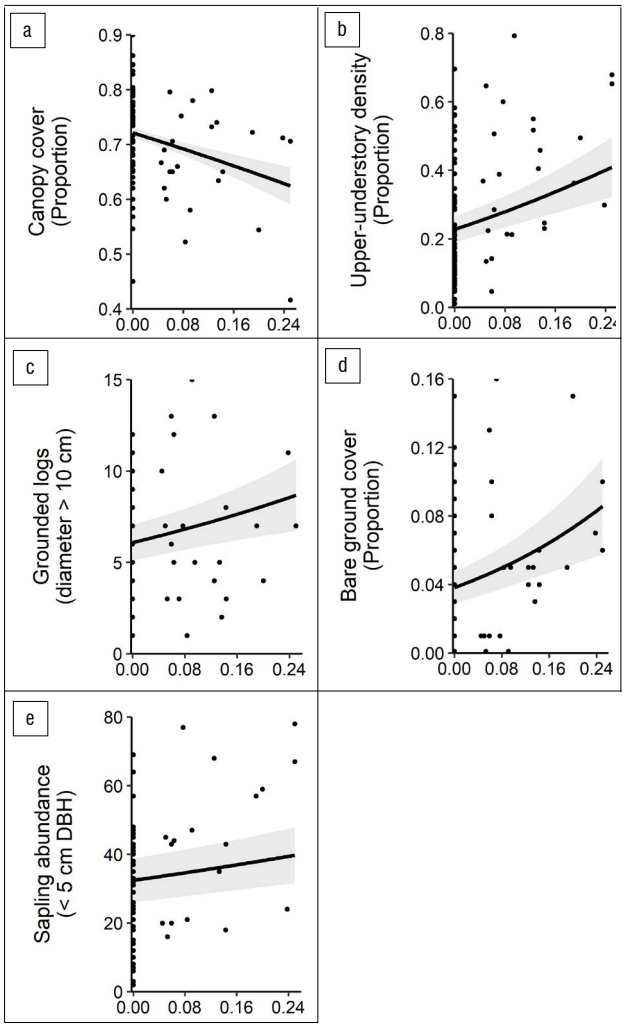


Figure 4: Effect of timber harvesting intensity, measured as the proportion of available stems (> 20 cm diameter at breast height/DBH) harvested, on (a) canopy cover, (b) upper-understorey layer (1–2 m) foliage density, (c) number of grounded logs, (d) bare ground cover and (e) sapling abundance. Relationships shown are derived from mixed models with forest included as a random effect (Table 1); however, graphic representations depict population-level predictions (i.e. excluding random effects).

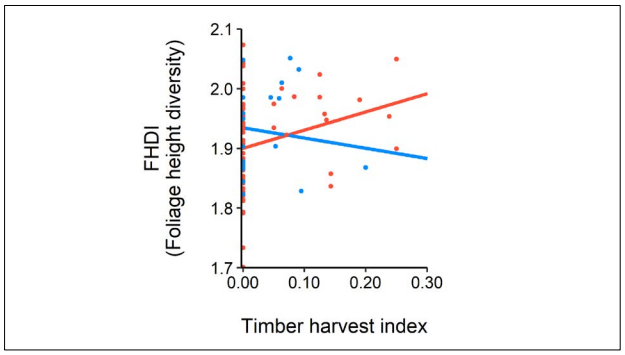


Figure 5: Effect of the two-way interaction between pole and timber harvesting on the foliage height diversity index (FHDI). The two lines represent variation in FHDI response to timber harvest intensity in the presence of low (i.e. 5%; red) and high (i.e. 20%; blue) intensities of pole harvesting, respectively. Relationships shown are derived from mixed models with forest included as a random effect (Table 1); however, graphic representations depict population-level predictions (i.e. excluding random effects).

Discussion

The findings of this study show that unregulated harvesting of medicinal bark, poles and timber results in multiple structural modifications to forest habitats in state forests of the Eastern Cape. Specifically, bark and timber harvesting created canopy gaps, while pole harvesting created understorey gaps, with variable implications for ground and understorey layer microhabitat structure, respectively. Our findings are thus in agreement with those of previous studies which have shown significant impacts of resource use on forest habitat structure in South Africa.[7,17,34] While the long-term ecological effects of harvest-mediated habitat modification are largely unknown[34], they represent changes to the natural disturbance regime, and are thus likely to have ramifying effects on forest patterns and processes[35], and faunal populations[26]. However, results of this study show that the extent of habitat modification is dependent on the nature and intensity of harvesting, and that different harvest activities, where occurring together at a fine spatial scale, may have interactive effects on habitat structure.

Bark harvesting

While several studies have examined the ecological implications of bark harvesting at individual and population levels[36-39], concurrent impacts on habitat structure have been relatively under-studied. Increasing bark harvesting intensities resulted in a decline in canopy cover and sapling abundance, and an increase in herb layer (0–0.5 m) and understorey layer (0–2 m) foliage density, grounded logs, and snag density (i.e. standing dead trees). These habitat modifications are the result of excessive bark removal from tree stem circumferences, preventing the transport of photosynthetic products to tree roots, leading to root loss or death, thereby driving declines in canopy health and potential tree mortality.[39,40] This creates gaps in the forest canopy, thereby increasing light availability to the forest floor such that ground and understorey layer foliage density increases. Over time, bark-harvested trees die and become snags which then decay, and dead branches drop to the ground, increasing the amount of grounded dead wood.

The substantial habitat-scale impacts of bark harvesting are perhaps best demonstrated by the close to 50% mean increase in snag abundance recorded across the four forests which experienced the highest levels of bark harvesting (Gomo, Manubi, Nqadu and Ntlaboya; Table 2), and associated increases in the number of grounded logs. While the ecological implications of the collection of dead wood for fuelwood from indigenous forests in South Africa have been cause for concern[6,41], with negative effects on cavity-nesting mammals and birds[42], few studies have highlighted the creation of dead wood in forests due to bark harvesting. The important ecological role of dead wood has long been recognised by ecologists.[43] However, the value to forest taxa of harvest-mediated dead wood creation, at the cost of living canopy trees of a select few species and canopy cover, is currently unknown but likely to be multifaceted and taxon dependent.

Pole harvesting

Unlike bark and timber harvesting, pole harvesting did not affect the forest canopy, but resulted in a decline in basal area and stem density of trees and saplings. This finding reflects the nature of pole harvesting wherein multiple understorey trees are harvested at a fine spatial scale, thereby creating gaps in the understorey, as shown by Boudreau and Lawes[17]. Despite the lack of any major canopy disturbances, as pole harvesting intensity increased, multiple understorey layer features were affected: foliage density at the herb layer (0–0.5 m) and herb cover increased, while leaf litter cover declined. Thus, while declines in basal area, tree and sapling density were a direct effect of harvesting, altered understorey and ground layer conditions are likely to be an indirect response driven by increases in light availability and soil moisture content due to a reduction in tree density.[17,44]

Although beyond the scope of this study, increased herb cover in understorey gaps caused by pole harvesting may supress seedling establishment.[44] Thus pole harvesting has the potential to alter not only structural habitat features, but also seedling recruitment, and therefore the maintenance of forest tree diversity. As indicated, changes

Table 2: Extent of harvesting activities recorded at each study forest, showing means (±s.e.) within each forest, with superscript letters indicating significant differences between forests

Harvest type	Lowland forests			Montane forests			Test statistic and *p*-value
	Manubi	Mqaba	Ntlaboya	Gomo	Nqadu	Pirie	
Bark harvest index	0.25 ± 0.08^{ab}	0.03 ± 0.01^{bc}	0.28 ± 0.11^{abc}	0.31 ± 0.07^{a}	0.16 ± 0.05^{abc}	0.01 ± 0.01^{c}	$\chi^2 = 27.02$, d.f. = 5, $p<0.01$
Proportion of pole-sized trees harvested per plot	0.10 ± 0.03^{a}	0.07 ± 0.02^{a}	0.05 ± 0.03^{ab}	0.14 ± 0.04^{a}	0.08 ± 0.03^{ab}	0.00 ± 0.00^{b}	$\chi^2 = 28.45$, d.f. = 5, $p<0.01$
Proportion of timber-sized trees harvested	0.07 ± 0.02^{a}	0.07 ± 0.02^{a}	0.02 ± 0.02^{ab}	0.02 ± 0.01^{ab}	0.00 ± 0.00^{b}	0.00 ± 0.00^{b}	$\chi^2 = 21.72$, d.f. = 5, $p<0.01$

in understorey conditions are dependent on the harvest intensity. Similarly, changes in seedling recruitment caused by pole harvesting are determined by understorey gap size,[44] with larger gaps causing a potential successional shift in seedling recruitment. However, Boudreau and Lawes[17] showed that under low harvesting intensity (11.6% of available pole-sized stems), pole harvesting did not negatively affect the long-term maintenance of tree diversity, suggesting that rates of pole harvesting measured in the current study (regional average of 7% of available pole-sized stems; Table 2) may not adversely affect tree species composition. However, modifications to understorey layer conditions may affect forest fauna. For example, leaf litter cover is a critical habitat for many forest invertebrates.[45]

Timber harvesting

At the habitat scale, timber harvesting resulted in canopy gaps through the selective extraction of canopy trees, driving an increase in upper-understorey (1–2 m) foliage density, sapling abundance, and bare ground cover. Furthermore, timber harvesting increased the number of grounded logs as a result of the large portions of harvested trees that are left in the forest after only the main stem of the harvested tree is removed. Furthermore, increases in dead wood may be associated with incidental tree damage associated with canopy-tree felling. Similar structural responses to selective timber harvesting have been shown by studies in tropical forests.[21,22,46] The creation of canopy gaps in forest systems represents a vital component of natural forest disturbance regimes, given their important role in promoting regeneration, tree diversity and habitat heterogeneity.[22,47-50] The gap phase represents a time of rapid plant growth[49], attributed to increased resource availability and/or decreased resource competition[51], demonstrated in the current study by the increased foliage density in the understorey. Furthermore, habitat conditions in canopy gaps compared to those in intact forest have been shown to differ significantly with respect to microclimate[51,52], detritus[27], productivity[53], and plant species composition[47,49]. Consequently, multiple forest taxa, including birds, reptiles and invertebrates, have been shown to distinguish between canopy gap and intact habitats.[22,46,50,54,55] This finding suggests that timber harvest activities, and concomitant habitat modifications, are likely to have ramifying effects on forest biodiversity.

The degree to which timber harvest activities affect forest biodiversity, beyond direct population-level impacts on target species, is likely to be dependent on the frequency of the disturbance, and the extent of incidental habitat damage. With regard to the former, selective harvesting practices in the Eastern Cape are likely to be less destructive than mechanised selective logging operations, which cause considerable damage through clearing for roads and log storage sites.[46] Informal timber harvesting in the Eastern Cape is generally un-mechanised, with felled timber split in the forest, and carried out on foot along narrow footpaths (J.L. personal observation). The frequency of disturbance is thus likely to be more cause for concern, as the harvest-driven increase in the proportion of forest-under-gap conditions is likely to have implications for ecosystem functioning.[35]

Interacting harvest effects

The positive relationship between foliage height diversity and biodiversity has been well established, and is based on niche theory which predicts that a greater diversity of habitats supports a greater diversity of species.[56] The decline in foliage height diversity in response to the interaction between pole and timber harvesting activities shown in this study indicates that, where these harvest types occur together at high rates, structural habitat complexity is reduced, and likely to negatively affect biodiversity at the habitat scale. This finding suggests that management strategies should limit the extent to which pole and timber harvesting activities occur together, and reduce the damage/lopping of smaller, non-target trees often associated with timber harvesting activities (J.L. personal observation), so as to maintain habitat heterogeneity in harvested areas.

Conclusion

The findings of this study indicate that resource use from state forests in the Eastern Cape has a significant impact on forest structure, although the nature and extent of the impact is dependent on the type and intensity of resource use. These results should be viewed within the context of forests that have a long history of human exploitation, from extensive colonial era logging to current subsistence and informal commercial harvesting of multiple forest products. However, the effects of long-term human exploitation are likely to have affected the current condition of all sampled forests, such that the findings of this study are indicative of habitat responses to more recent resource use disturbances. Similarly, while habitat structure is modified by random natural disturbances, such as windfalls, lightning or fire-spotting, which are vital components of natural disturbance–recovery regimes that maintain forest dynamics, resource use represents disturbances that occur in addition to these natural disturbances under which forest species have adapted, and thus may affect ecosystem persistence and resilience. Further research is needed to determine specific levels of resource use that can be sustained without negatively affecting forest biodiversity. Specifically, research regarding the impact of resource use on forest taxa at multiple trophic levels is needed to provide insight into ecosystem-wide implications of harvest-mediated habitat modification, and to contribute to the development of ecologically informed forest management policies.

Acknowledgements

We thank Welile Kedama of the South African Department of Environment, Forestry and Fisheries for granting permission to conduct research within state forests of the Eastern Cape, and to all Forestry personnel who assisted with fieldwork. C.J. Geldenhuys is thanked for meaningful discussion that helped guide this study. We are grateful to A. Wannenburgh for providing input on the first draft of this manuscript, and for the study site map. R. Duker is thanked for assistance in developing the data collection methodology. An anonymous reviewer is thanked for comments on the first draft of this manuscript. James Pryke is thanked for his comments on the thesis chapter from which this manuscript was derived, which have improved the manuscript.

Competing interests

We declare that there are no competing interests.

Data availability

The data sets generated and/or analysed during the current study are available from the corresponding author on request.

Authors' contributions

J.L. designed the study methodology, conducted data collection and data analysis, and wrote the initial draft of the manuscript. M.I.C. conceptualised the study, acquired funding, and contributed extensively to the writing of the manuscript revision.

References

1. Arroyo-Rodríguez V, Melo FPL, Martínez-Ramos M, Bongers F, Chazdon RL, Meave JA, et al. Multiple successional pathways in human-modified tropical landscapes: New insights from forest succession, forest fragmentation and landscape ecology research: Multiple successional pathways. Biol Rev. 2017;92:326–340. https://doi.org/10.1111/brv.12231
2. Cooper TJG, Wannenburgh AM, Cherry MI. Atlas data indicate forest dependent bird species declines in South Africa. Bird Conserv Int. 2017;27:337–354. https://doi.org/10.1017/S095927091600040X
3. Shackleton CM, Mograbi PJ, Drimie S, Fay D, Hebinck P, Hoffman MT, et al. Deactivation of field cultivation in communal areas of South Africa: Patterns, drivers and socio-economic and ecological consequences. Land Use Policy. 2019;82:686–699. https://doi.org/10.1016/j.landusepol.2019.01.009
4. Bond WJ, Midgley GF. A proposed CO_2-controlled mechanism of woody plant invasion in grasslands and savannas. Glob Chang Biol. 2000;6:865–869. https://doi.org/10.1046/j.1365-2486.2000.00365.x
5. Berliner DD. Systematic conservation planning for South Africa's forest biome: An assessment of the conservation status of South Africa's forests and recommendations for their conservation [PhD thesis]. Cape Town: University of Cape Town; 2009.
6. Castley JG, Kerley GIH. The paradox of forest conservation in South Africa. For Ecol Manage. 1996;85:35–46. https://doi.org/10.1016/S0378-1127(96)03748-6
7. Hoppe-Speer SCL, Adams JB, Bailey D. Present state of mangrove forests along the Eastern Cape coast, South Africa. Wetl Ecol Manag. 2105;23:371–383. https://doi.org/10.1007/s11273-014-9387-x
8. Obiri J, Lawes M, Mukolwe M. The dynamics and sustainable use of high-value tree species of the coastal Pondoland forests of the Eastern Cape Province, South Africa. For Ecol Manage. 2002;166:131–148. https://doi.org/10.1016/S0378-1127(01)00665-X
9. Von Maltitz G, Mucina L, Geldenhuys C, Lawes MJ, Eeley HAC, Adie H, et al. Classification system for South African indigenous forests: An objective classification for the Department of Water Affairs and Forestry. Pretoria: CSIR; 2003.
10. King NL. The exploitation of indigenous forests in South Africa. S Afr J Bot. 1941;48:455–480. https://doi.org/10.1080/03759873.1941.9631098
11. Paumgarten F, Shackleton CM. The role of non-timber forest products in household coping strategies in South Africa: The influence of household wealth and gender. Popul Environ. 2011;33:108–131. https://www.jstor.org/stable/41487565
12. Dold AP, Cocks ML. The trade in medicinal plants in the Eastern Cape Province, South Africa. S Afr J Sci. 2002;98:589–597.
13. Shackleton C, Shackleton S. The importance of non-timber forest products in rural livelihood security and as safety nets: A review of evidence from South Africa. S Afr J Sci. 2004;100:658–664.
14. Geldenhuys CJ. Bark harvesting for traditional medicine: from illegal resource degradation to participatory management. Scand J For Res. 2004;19:103–115. https://doi.org/10.1080/14004080410034182
15. Obiri JAF, Lawes MJ. Attitudes of coastal-forest users in Eastern Cape Province to management options arising from new South African forest policies. Environ Conserv. 2002;29:519–529. https://doi.org/10.1017/S0376892902000371
16. Von Maltitz GP, Shackleton S. Use and management of forests and woodlands in South Africa: Stakeholders, institutions and processes from past to present. In: Lawes MJ, Eeley HAC, Shackleton CM, Geach BGS, editors. Indigenous forests and woodlands in South Africa: Policy, people and practice. Pietermaritzburg: University of KwaZulu-Natal Press; 2004. p. 109–112.
17. Boudreau S, Lawes MJ. Small understorey gaps created by subsistence harvesters do not adversely affect the maintenance of tree diversity in a sub-tropical forest. Biol Conserv. 2005;126:279–286. https://doi.org/10.1016/j.biocon.2005.06.004
18. Ticktin T. The ecological implications of harvesting non-timber forest products: Ecological implications of non-timber harvesting. J Appl Ecol. 2004;41:11–21. https://doi.org/10.1111/j.1365-2664.2004.00859.x
19. Belsky AJ, Blumenthal DM. Effects of livestock grazing on stand dynamics and soils in upland forests of the interior west. Conserv Biol. 1997;11:315–327. https://doi.org/10.1046/j.1523-1739.1997.95405.x
20. Yates CJ, Norton DA, Hobbs RJ. Grazing effects on plant cover, soil and microclimate in fragmented woodlands in south-western Australia: Implications for restoration. Austral Ecol. 2000;25:36–47. https://doi.org/10.1111/j.1442-9993.2000.tb00005.x
21. Sekercioglu CH. Effects of forestry practices on vegetation structure and bird community of Kibale National Park, Uganda. Biol Conserv. 2002;107:229–240. https://doi.org/10.1016/s0006-3207(02)00097-6
22. Wunderle JM, Henriques LMP, Willig MR. Short-term responses of birds to forest gaps and understory: an assessment of reduced-impact logging in a lowland Amazon forest. Biotropica. 2006;38:235–255. https://doi.org/10.1111/j.1744-7429.2006.00138.x
23. Fashing PJ. Mortality trends in the African cherry (*Prunus africana*) and the implications for colobus monkeys (*Colobus guereza*) in Kakamega Forest, Kenya. Biol Conserv. 2004;120:449–459. https://doi.org/10.1016/j.biocon.2004.03.018
24. Leaver J, Mulvaney J, Ehlers Smith DA, Ehlers Smith YC, Cherry MI. Response of bird functional diversity to forest product harvesting in the Eastern Cape, South Africa. For Ecol Manage. 2019;445:82–95. https://doi.org/10.1016/j.foreco.2019.04.054
25. Asefa A, Davies AB, McKechnie AE, Kinahan AA, Van Rensburg BJ. Effects of anthropogenic disturbance on bird diversity in Ethiopian montane forests. Condor. 2017;119:416–430. https://doi.org/10.1650/CONDOR-16-81.1
26. Bawa KS, Seidler R. Natural forest management and conservation of biodiversity in tropical forests. Conserv Biol. 1998;12:46–55. https://doi.org/10.1111/j.1523-1739.1998.96480.x
27. Cleary DFR, Boyle TJB, Setyawati T, Anggraeni CD, Loon EEV, Menken SBJ. Bird species and traits associated with logged and unlogged forest in Borneo. Ecol Appl. 2017;17:1184–1197. https://doi.org/10.1890/05-0878
28. Gardner CJ, Jasper LD, Eonintsoa C, Duchene JJ, Davies ZG. The impact of natural resource use on bird and reptile communities within multiple-use protected areas: Evidence from sub-arid Southern Madagascar. Biodivers Conserv. 2016;25:1773–1793. https://doi.org/10.1007/s10531-016-1160-4
29. Presley SJ, Willig MR, Wunderle JM, Saldanha LN. Effects of reduced-impact logging and forest physiognomy on bat populations of lowland Amazonian forest: Bat responses to logging in Amazonia. J Appl Ecol. 2007;45:14–25. https://doi.org/10.1111/j.1365-2664.2007.01373.x
30. Menon T, Sridhar H, Shahabuddin G. Effects of extractive use on forest birds in Western Himalayas: Role of local and landscape factors. For Ecol Manage. 2019;448:457–465. https://doi.org/10.1016/j.foreco.2019.06.033
31. Sekercioglu CH. Increasing awareness of avian ecological function. Trends Ecol Evol. 2006;21:464–471. https://doi.org/10.1016/j.tree.2006.05.007
32. Byrne LB. Habitat structure: A fundamental concept and framework for urban soil ecology. Urban Ecosyst. 2007;10:255–274. https://doi.org/10.1007/s11252-007-0027-6
33. Cunningham AB. African medicinal plants: Setting priorities at the interface between conservation and primary health care. People and Plants Working Paper 1. Paris: UNESCO; 1993.
34. Lawes MJ, Griffiths ME, Boudreau S. Colonial logging and recent subsistence harvesting affect the composition and physiognomy of a podocarp dominated Afrotemperate forest. For Ecol Manage. 2007;247:48–60. https://doi.org/10.1016/j.foreco.2007.04.012

35. Royo AA, Carson WP. On the formation of dense understory layers in forests worldwide: Consequences and implications for forest dynamics, biodiversity, and succession. Can J For Res. 2006;36:1345–1362. https://doi.org/10.1139/x06-025

36. Chungu D, Muimba-Kankolongo A, Roux J, Malambo F. Bark removal for medicinal use predisposes indigenous forest trees to wood degradation in Zambia. Southern Forests. 2007;69:157–163. https://doi.org/10.2989/SHFJ.2007.69.3.4.354

37. Gaoue OG, Ticktin T. Patterns of harvesting foliage and bark from the multipurpose tree *Khaya senegalensis* in Benin: Variation across ecological regions and its impacts on population structure. Biol Conserv. 2007;137:424–436. https://doi.org/10.1016/j.biocon.2007.02.020

38. Guedje NM, Zuidema PA, During H, Foahom B, Lejoly J. Tree bark as a non-timber forest product: The effect of bark collection on population structure and dynamics of *Garcinia lucida* Vesque. For Ecol Manage. 2007;240:1–12. https://doi.org/10.1016/j.foreco.2006.09.029

39. Vermeulen W, Geldenhuys C, Esler K. Response of *Ocotea bullata*, *Curtisia dentata* and *Rapanea melanophloeos* to medicinal bark stripping in the southern Cape, South Africa: Implications for sustainable use. Southern Forests. 2012;74:183–193. https://doi.org/10.2989/20702620.2012.717384

40. Cunningham AB. Applied ethnobotany: People, wild plant use and conservation. London: Earthscan Publications; 2001.

41. Berry MG, Robertson BL, Campbell EE, Bredenkamp GJ. Impact of cutting and collecting of firewood associated with informal settlement in the south-eastern Cape coastal zone. S Afr J Bot. 2005;71:179–190. https://doi.org/10.1016/S0254-6299(15)30131-9

42. Du Plessis MA. The effects of fuelwood removal on the diversity of some cavity-using birds and mammals in South Africa. Biol Conserv. 1995;74:77–82. https://doi.org/10.1016/0006-3207(95)00016-w

43. Zhou L, Dai L, Gu H, Zhong L. Review on the decomposition and influence factors of coarse woody debris in forest ecosystem. J For Res. 2007;18:48–54. https://doi.org/10.1007/s11676-007-0009-9

44. Louw SL. The effect of the spatial scale of tree harvesting on woody seedling establishment and tree dynamics at Ongoye Forest Reserve. Pietermaritzburg: University of KwaZulu-Natal; 2010.

45. Banks JE, Jackson C, Hannon LM, Thomas CM, Baya A, Njoroge L. The cascading effects of elephant presence/absence on arthropods and an Afrotropical thrush in Arabuko-Sokoke Forest, Kenya. Afr J Ecol. 2010;48:1030–1038. https://doi.org/10.1111/j.1365-2028.2010.01211.x

46. Thiollay JM. Influence of selective logging on bird species diversity in a Guianan rain forest. Conserv Biol. 1992;6:47–63. https://doi.org/10.1016/0006-3207(93)90360-d

47. Kneeshaw DD, Bergeron Y. Canopy gap characteristics and tree replacement in the southeastern boreal forest. Ecology. 1998;79:783–794. https://doi.org/10.2307/176578

48. Obiri JAF, Lawes MJ. Chance versus determinism in canopy gap regeneration in coastal scarp forest in South Africa. J Veg Sci. 2004;15:539–547. https://doi.org/10.1111/j.1654-1103.2004.tb02293.x

49. Brokaw NVL. Gap-phase regeneration in a tropical forest. Ecology. 1985;66:682–687. https://doi.org/10.2307/1940529

50. Levey DJ. Tropical wet forest treefall gaps and distributions of understory birds and plants. Ecology. 1988;69:1076–1089. https://doi.org/10.2307/1941263

51. Canham CD. An index for understory light levels in and around canopy gaps. Ecology. 1988;69:1634–1638. https://doi.org/10.2307/1941664

52. Gray AN, Spies TA, Easter MJ. Microclimatic and soil moisture responses to gap formation in coastal Douglas-fir forests. Can J For Res. 2002;32:332–343. https://doi.org/10.1139/x01-200

53. Prescott CE. The influence of the forest canopy on nutrient cycling. Tree Physiol. 2002;22:1193–1200. https://doi.org/10.1093/treephys/22.15-16.1193

54. Richards LA, Windsor DM. Seasonal variation of arthropod abundance in gaps and the understorey of a lowland moist forest in Panama. J Trop Ecol. 2007;23:169–176. https://doi.org/10.1017/S0266467406003907

55. Vitt LJ, Avila-Pires TCS, Caldwell JP, Oliveira VRL. The impact of individual tree harvesting on thermal environments of lizards in Amazonian rain forest. Conserv Biol. 1998;12:654–664. https://doi.org/10.1111/j.1523-1739.1998.96407.x

56. MacArthur RH, MacArthur JW. On bird species diversity. Ecology. 1961;42:594–598. https://doi.org/10.2307/1932254

AUTHORS:
Michiel M. Scholtz[1,2]
Frederick W.C. Neser[2]
Mahlako L. Makgahlela[1,2]

AFFILIATIONS:
[1]Animal Production, Agricultural Research Council, Pretoria, South Africa
[2]Department of Animal, Wildlife and Grassland Sciences, University of the Free State, Bloemfontein, South Africa

CORRESPONDENCE TO:
Michiel Scholtz

EMAIL:
GScholtz@arc.agric.za

DATES:
Received: 20 April 2020
Revised: 26 June 2020
Accepted: 29 June 2020
Published: 29 Sep. 2020

HOW TO CITE:
Scholtz MM, Neser FWC, Makgahlela ML. A balanced perspective on the importance of extensive ruminant production for human nutrition and livelihoods and its contribution to greenhouse gas emissions. S Afr J Sci. 2020;116(9/10), Art. #8192, 3 pages. https://doi.org/10.17159/sajs.2020/8192

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:
animal source foods, cognitive development, livestock, greenhouse gases, nutrition

FUNDING:
None

A balanced perspective on the importance of extensive ruminant production for human nutrition and livelihoods and its contribution to greenhouse gas emissions

There is a general perception that ruminants produce large quantities of greenhouse gases which contribute to global warming. Sometimes percentages are quoted out of context. For example, the percentage quoted for developed countries indicates the greenhouse gas contribution from livestock is less than 6%, while that for developing countries is 40–50%. However, the reason for this relatively low contribution from developed countries is because of very high contributions from other sectors. Ruminant production also is in the spotlight as it is the world's largest user of land and South Africa is no exception. Only ruminants can utilise areas of non-arable land where the vegetation is rich in fibre and convert this fibre into high-quality nutrients for human consumption. Foods from animal sources (including ruminants) are essential for the human diet, as they support early childhood and cognitive development. Many rural households depend on ruminants and these animals are central to the livelihoods and well-being of these communities. The negative effects of red meat on human health and the negative environmental impact of livestock production are overemphasised, while the higher bioavailability of nutrients from livestock source foods, which stimulates mental and cognitive development compared to vegetarian or grain based foods, is ignored. Here we estimate that livestock are responsible for only 4% of the world's greenhouse gases through methane production. We also highlight that if the high fibre vegetation is not utilised by livestock, it will still produce greenhouse gases through burning or rotting, without any benefit to humans. Livestock source foods are important if global nutritional, educational and economic needs are to be met; and this message should be conveyed to the public.

Significance:

- We propose that a balanced message should be conveyed to the broader scientific community and the public on the role of livestock in meeting global nutritional needs and contributing to global warming.
- Livestock source foods are important if the global nutritional, educational and economic needs are to be met and can be used to feed developing countries out of poverty.

Livestock is often blamed for producing large quantities of greenhouse gases (GHG) and thus contributing to global warming. This general perception, that livestock makes a huge contribution to global warming, follows on the Food and Agriculture Organization's publication of *Livestock's Long Shadow*[1], in which it is reported that livestock is responsible for 18% of the total greenhouse gas (GHG) emissions, although it was later scaled down to 14.5% of the total anthropogenic GHG emissions[2]. Furthermore, these percentages are at times irresponsibly quoted out of context. For example, the percentages quoted for developed countries indicate that the GHG contribution from agriculture is less than 6%. However, the reason for this low contribution is a result of the very high contributions from the energy, manufacturing, mining and other sectors. Developing countries have lower energy consumption and smaller manufacturing and mining sectors. Thus, although the relative contribution from agriculture ranges between 40% and 50%, the actual contribution is still lower than 6%.[3]

Livestock are characterised into ruminants (cattle, sheep, goats and water buffalo) and monogastric animals (pigs, poultry, donkeys and horses). Ruminants generate a large amount of GHGs, mainly in the form of methane (CH_4), through enteric fermentation, which is a natural by-product of anaerobic microbial fermentation and manure storage. The methane produced in the rumen is emitted by belching. It is estimated that about 80% of the GHG from livestock comes from ruminants and that they are responsible for more than 90% of the total CH_4 emissions from livestock.[4] Thus, while ruminants play an important role in providing high-quality protein essential for human diets, they are also an important source of animal GHG emissions.

That ruminant production is in the spotlight is understandable, as it is the world's largest individually identifiable producer of GHGs and user of land, and South Africa is a clear example of the latter. About 84% of South Africa's land is available for agriculture, but most of this land cannot be used for crop production. Only 13% of South Africa's land area is arable, with the greater part (71%) only suitable for extensive ruminant production.[3]

Ruminants are often the scapegoat for GHG emissions, with their strategic value in the supply of human nutrition often disregarded. In many parts of sub-Saharan Africa ruminants play an important role in human diets. Most of the fibre rich vegetation on land can only be utilised by ruminants which convert it into high-quality protein and other nutrients for human consumption.[3]

The importance of animal source foods should be recognised, because any reduction in the consumption of ruminant source foods (meat and dairy products) will compromise the intake of the nutrients that are supplied in relatively large proportions in meat and dairy products. In cases where there is evidence of low nutritional

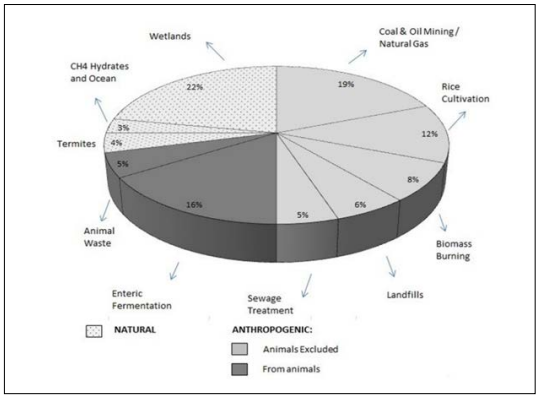
status, the risk will be bigger.[5] Millions of children in developing countries already suffer from impaired cognitive development as a result of poor nutrition due to the insufficient consumption of livestock source foods.[6] A relationship between early and even later cognitive development and the consumption of livestock source foods by pregnant and lactating women, and children, has been established.[7] The reason for this is that livestock source foods supply nutrients with a higher bioavailability that stimulates mental and cognitive development, compared to vegetarian- or grain-based foods.[6] It should also be noted that immunoglobulin, which is present in cow's milk, assists the human body in its fight against bacteria, viruses and allergens.[8] Cow's milk may therefore influence the development of the immune system in the upper respiratory tract[9], which may play a role in the body's defence against the novel coronavirus disease (COVID-19).

It should also be noted that small-scale and subsistence farmers keep ruminants for a variety of purposes. Many rural households are dependent on ruminants for milk, meat, hides, horns, fertiliser, draught and income.[10,11] Ruminants are therefore central to the livelihoods and well-being of these households.

In spite of the importance of livestock production, the recently published report by the EAT-Lancet Commission[12] downplays the significant role of livestock in providing the valued nutritional elements. Instead the report focuses on the negative effects of red meat source foods (from ruminants) on human health and overestimates the negative environmental impact of livestock production. It is not only the EAT-Lancet Commission that is overplaying the effect of methane emissions from ruminant livestock on global warming. This is also done by many other groups with their own agenda, such as the Meat-Free Monday campaign. Capper[13] indicates that if all the USA's 313 million inhabitants adopted meat-free Mondays, the annual reduction in GHG emissions is estimated to be only 0.3%. The information used by these advocacy groups is often overly simplistic, and based on misconceptions, a lack of knowledge and incorrect calculations.[13-15]

Climate change – the increase in average temperature and an associated increase in the frequency of extreme weather events – and global warming are consequences of an increased production of carbon dioxide (CO_2), methane (CH_4), nitrous oxide (N_2O), hydrofluorocarbons, perfluorocarbons and sulfur hexafluoride (SF_6), collectively referred to as GHG.[16,17]

Of these six GHGs, the most important are CO_2, CH_4 and N_2O because they are closely associated with human activities.[18] Carbon dioxide is produced through burning fossil fuels and other biological material (e.g. solid waste and trees), as well as from certain chemical reactions, whereas the CO_2 from livestock originates from the aerobic digestion of manure and other waste. It is important to note that CO_2 is removed from the atmosphere through the process of sequestration, in which plants absorb CO_2 as part of the biological carbon cycle, whereas nitrous oxide is emitted through the combustion of fossil fuels and other waste products, agricultural and industrial activities and during the treatment of waste water.[3,17] The natural and anthropogenic sources of CH_4 production are illustrated in Figure 1.



Source: *Adapted from Augenbraun et al.*[22]

Figure 1: Natural and anthropogenic sources of methane

The major GHGs related to livestock production are CO_2, CH_4 and N_2O. Their characteristics in terms of atmospheric concentration, atmospheric lifetime and heating potential are summarised in Table 1.

Table 1: Major greenhouse gases (GHG) related to lifestyle production and their characteristics

GHG	Carbon dioxide	Methane	Nitrous oxide
Atmospheric concentration (%)	49	18	6
Atmospheric lifetime (years)	100–200	12	114
Heating potential (CO_2-eq)	1	23	296

Sources: *Adapted from Clark et al.*[18]*, IPCC*[21] *and Biotech*[23]

Most of the CH_4 produced by livestock originates from enteric fermentation by ruminants. It is therefore important to identify the sources of CH_4 production as indicated in Figure 1. Livestock and rice are the two most important single food sources for the developing world. However, these two food sources are also responsible for the production of large quantities of anthropogenic CH_4. Livestock are responsible for 21% (16% from enteric fermentation and 5% from animal waste, including manure) and rice cultivation for 12% of anthropogenic CH_4 (Figure 1).

A simple calculation can be made using the information from Table 1 and Figure 1. Livestock contributes 21% of anthropogenic CH_4 production and the atmospheric concentration of CH_4 forms only 18% of the GHG emissions. Thus $0.21 \times 0.18 = 0.04$. This implies that livestock is responsible for only 4% of the world's GHG emissions through CH_4 production.

It should be noted that the global warming potential of CH_4 is approximately 23 times more than that of CO_2, but its atmospheric lifetime is 12 years compared to 100–200 years for CO_2 (Table 1). Although CH_4 has a larger effect, the duration of the effect is much shorter. This is one of the aspects that is frequently ignored. In addition, a large percentage of ruminant production is in developing countries, and supports rural livelihoods[3,19], and, in many cases, is a more environmentally friendly method of producing ruminant source foods[20].

An important question to ask is: What will happen with the vegetation if it is not consumed by product-producing (meat, milk and fibre) ruminants? There are three possible consequences for the vegetation: (1) it can be consumed by wild animals that will also emit CH_4, (2) it can burn during veldfires/wildfires, which will produce CO_2 that is released into the atmosphere with an atmospheric lifetime of 100–200 years, or (3) it can rot and produce N_2O with a global warming potential of almost 300 times more than that of CO_2. It is also important to note that the domesticated ruminants to a large extent replaced wildlife on the same land and the wildlife inhabiting the area before also produced CH_4. The nett effect may thus be more or less the same. During the regrowth of the vegetation grazed by ruminants, CO_2 is absorbed from the atmosphere. This carbon sequestration that occurs naturally has been neglected and therefore the quantitative effect thereof is not known.

It should also be noted that differences in livestock production systems between countries can influence the carbon footprint of livestock products. This is especially the case between developed countries in Europe and the Americas versus developing countries. Some of these differences relate to production systems (intensive versus extensive), manure storage and the application thereof, as well as feed production, transport and processing.[21] Many of the current methods used to estimate carbon footprints are based on generic assumptions that do not take into account the different production systems. The principle of carbon sequestration during the regrowth of the vegetation is sometimes also ignored.

The bottom line is that livestock is important for human sustainability, as it plays a critical role in increasing food security, improving nutrition, reducing poverty, and improving human health.[6] This important contribution of livestock source foods to humans should be considered in all sustainability and climate change debates. It is therefore crucial that a balanced message be conveyed to the broader scientific community as well as the public on the role of livestock in meeting global nutritional needs and in contributing to global warming. Livestock source foods are important for global nutritional, educational and economic needs to be met and can feed developing countries out of poverty.

Acknowledgements

This Research Letter was stimulated by the project titled 'Sustainable natural resource utilisation and livestock production in the era of climate change' which is supported by the ARC Climate Change Collaboration Centre. Mrs Georgette Pyoos-Daniels is acknowledged for assistance with the compilation of the references.

Competing interests

We declare that there are no competing interests.

Authors' contributions

M.M.S. conceived and conceptualised the Research Letter and also wrote the initial manuscript. Both F.W.C.N. and M.L.M. made valuable proposals on improving the manuscript. All authors participated in writing the final manuscript and approved it.

References

1. Steinfeld H, Gerber P, Wassenaar T, Castel V, Rosales M, De Haan C. Livestock's long shadow: Environmental issues and options. Rome: FAO; 2006. https://doi.org/10.1186/2041-7136-4-1
2. Gerber P, Steinfeld H, Henderson B, Mottet A, Opio C, Dijkma J, et al. Tackling climate change through livestock: A global assessment of emissions and mitigation opportunities [document on the Internet]. c2013 [cited 2020 Jan 30]. Available from: http://www.fao.org/3/a-i3437e.pdf
3. Scholtz MM, Van Ryssen JBJ, Meissner HH, Laker MC. A South African perspective on livestock production in relation to greenhouse gases and water usage. S Afr J Anim Sci. 2013;43(3):247–254. https://doi.org/10.4314/sajas.v43i3.2
4. Opio C, Gerber P, Mottet A, Falcucci A, Tempio G, MacLeod M, et al. Greenhouse gas emissions from ruminant supply chains – A global life cycle assessment [document on the Internet]. c2013 [cited 2020 Jan 30]. Available from: http://www.fao.org/3/i3461e/i3461e.pdf
5. Schönfeldt HC, Pretorius B, Hall N. 'Fish, chicken, lean meat and eggs can be eaten daily': A food-based dietary guideline for South Africa. S Afr J Clin Nutr. 2013;26(3)(suppl):S66–S76. https://doi.org/10.5772/67110
6. Adesogan AT, Havelaar AH, McKune SL, Eilittäd M, Dahla GE. Sustainability problem or malnutrition and sustainability solution? Perspective matters. Global Food Security. 2020;25, Art. #100325. https://doi.org/10.1016/j.gfs.2019.100325
7. Balehegn M, Mekuriaw Z, Miller L, McKune S, Adesogan AT. Animal-sourced foods for improved cognitive development. Anim Front. 2019;9(4):50–57. https://doi.org/10.1093/af/vfz039
8. Ulfman LH, Leusen JHW, Savelkoul HFJ, Warner JO, Van Neeren RJJ. Effects of bovine immunoglobins on immune function, allergy and infection. Front Nutr. 2018;5:52. https://doi.org/10.3389/fnut.2018.00052
9. Reynolds HY. Immunoglobin G and its function in the human respiratory tract. Mayo Clin Proc. 1988;63:161–174. http://doi.org/10.1016/s0025-6196(12)64949-0
10. Chimonyo M, Kusina NT, Hamudikuwanda H, Nyoni O. A survey on land use and usage of cattle for draught in a smallholder farming area of Zimbabwe. J Appl Sci South Afr. 1999;5:111–121. https://doi.org/10.4314/jassa.v5i2.16914
11. Dovie DBK, Shackleton CM, Witkowski ETF. Valuation of communal area livestock benefits, rural livelihoods and related policy issues. Land Use Policy. 2006;23:260–271. https://doi.org/10.1016/j.landusepol.2004.08.004
12. EAT Lancet Commission. The EAT-Lancet Commission on Food, Planet, Health: Can we feed a future population of 10 billion people a healthy diet within planetary boundaries? [document on the Internet]. c2019 [cited 2020 Mar 22]. Available from: https://eatforum.org/eat-lancet-commission/
13. Capper JL. Should we reject animal source foods to save the planet? A review of the sustainability of global livestock production. S Afr J Anim Sci. 2013;43:233–246. https://doi.org/10.4314/sajas.v43i3.1
14. Meissner HH, Scholtz MM, Engelbrecht FA. Sustainability of the South African livestock sector towards 2050. Part 2: Challenges, changes and required implementations. S Afr J Anim Sci. 2013;43:298–319. https://doi.org/10.4314/sajas.v43i3.6
15. Capper JL, Cady RA, Bauman DD. Demystifying the environmental sustainability of food production. In: Proceedings of the Cornell Nutrition Conference. Syracuse, NY: Cornell University; 2009. p. 187–203. https://doi.org/10.2527/jas.2009-1781
16. United States Environmental Protection Agency. Inventory of U.S. Greenhouse Gas Emissions and Sinks 1990 – 2018 [document on the internet]. c2018 [cited 2020 Apr 10]. Available from: https://www.epa.gov/ghgemissions/inventory-us-greenhouse-gas-emissions-and-sinks
17. United States Environmental Protection Agency. Overview of greenhouse gases [document on the Internet]. c2019 [cited 2020 Apr 25]. Available from: https://www.epa.gov/ghgemissions/overview-greenhouse-gases
18. Clark H, Klein C, Newton P. Potential management practices and technologies to reduce nitrous oxide, methane and carbon dioxide emissions from New Zealand agriculture. Ngaherehere, NZ: Ministry of Agriculture and Forestry; 2001 [cited 2020 Apr 03]. Available from: http://www.maf.govt.nz/mafnet/rural-nz/sustainable-resource-use/climate/green-house-gas-migration/ghg-mitigation.htm
19. Meissner HH, Scholtz MM, Palmer AR. Sustainability of the South African livestock sector towards 2050. Part 1: Worth and impact of the sector. S Afr J Anim Sci. 2013;43(3):282–297. https://doi.org/10.4314/sajas.v43i3.5
20. Scholtz MM, Du Toit J, Neser FWC. Antagonism in the carbon footprint between beef and dairy production systems. S Afr J Anim Sci. 2014;44:17–20. http://doi.org./10.4314/sajas.44i5.4
21. Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL, editors. Climate change 2007: The physical science basis contribution of Working Group I to the fourth assessment report of the Intergovernmental Panel on Climate Change. Cambridge, UK: Cambridge University Press; 2007. Available from: https://archive.ipcc.ch/pdf/assessment-report/ar4/wg1/ar4-wg1-frontmatter.pdf
22. Augenbraun H, Matthews E, Sarma D. The global methane cycle [document on the Internet]. c2010 [cited 2020 Apr 16]. Available from: http://icp.giss.nasa.gov/education/methane/intro/cycle.html
23. Biotech LTD. Agribusiness and Climate Change Seminar: Nutrition and low methane diet. Figures obtained from IPCC 1996 National Emissions [document on the Internet]. c1996 [cited 2020 May 25]. Available from: www.neembiotech.com

AUTHORS:
Sheunesu Ruwanza[1]
Edward S. Mhlongo[1]

AFFILIATION:
[1]Centre of Excellence for Invasion Biology, Department of Environmental Science, Rhodes University, Makhanda, South Africa

CORRESPONDENCE TO:
Sheunesu Ruwanza

EMAIL:
ruwanza@yahoo.com

DATES:
Received: 11 May 2020
Revised: 13 July 2020
Accepted: 22 July 2020
Published: 29 Sep. 2020

HOW TO CITE:
Ruwanza S, Mhlongo ES. *Lantana camara* invasion along road–river interchanges and roadsides in Soutpansberg, Vhembe Biosphere Reserve in South Africa. S Afr J Sci. 2020;116(9/10), Art. #8302, 5 pages. https://doi.org/10.17159/sajs.2020/8302

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:
disturbance, invasive plant, management, road survey

FUNDING:
DST-NRF Centre of Excellence for Invasion Biology; South African National Research Foundation (grant TTK190117408656)

Lantana camara invasion along road–river interchanges and roadsides in Soutpansberg, Vhembe Biosphere Reserve in South Africa

Roads and rivers act as conduits of alien plant invasion; however, little is known regarding the abundance and invasion extent of *Lantana camara*, an invasive shrub, along road–river interchanges and roadsides in South Africa. We assessed the effects of road–river interchanges and roadsides as invasion corridors that facilitate *L. camara* invasion. A road survey method was used to measure the invasion extent of *L. camara* along road–river interchanges and roadsides from national and regional roads covering 446 km in Soutpansberg, Vhembe Biosphere Reserve, South Africa. *L. camara* occurred along 21 of the 48 surveyed road–river interchanges and its abundance and cover were similar between road–river interchanges and roadsides, although height and diameter of *L. camara* were greater along road–river interchanges than roadsides. Other alien species that dominated road–river interchanges were *Solanum mauritianum, Caesalpinia decapetala* and *Rubus rigidus*. Our results indicate that *L. camara* dominates both road–river interchanges and roadsides, therefore roads and bridges should be considered important targets for *L. camara* control.

Significance:

- Despite the huge efforts by the South African government to control invasive alien plants, roads and rivers continue to act as important conduits for invasion and therefore there is a need to manage them.
- We propose that South African roads be considered important targets for efforts to control invasive alien plants.
- A policy and legislative framework around invasive alien plant removal during routine road maintenance operations is required.

Introduction

Roads and bridges function as corridors for alien plant invasion and spread, although their importance has been neglected in managing invasive plants.[1,2] Rahlao et al.[3] highlighted that habitat disturbances at roadsides and road–river interchanges promote the establishment of invasive alien plants through the creation of suitable conditions for plant invasion through water and nutrient resource availability. Indeed, habitat fragmentation along roads and bridges enhances the establishment of disturbance-tolerant invasive plants, whilst withholding the establishment of local species.[3,4] Certainly, the colonisation of roadsides by invasive alien plants may act as a source for ensuing invasion into the neighbouring natural communities.[3] Few studies in South Africa have interrogated the effects of roadsides and road–river interchanges on invasion abundance and extent,[3,4] yet roads and rivers act as conduits for the efficient dispersal of alien plants. Against this background, we assessed the effects of road–river interchanges and roadsides as invasion corridors that facilitate *Lantana camara* invasion.

Lantana camara is an ornamental shrub which originates from South America. It has been declared a category 1b invasive plant in South Africa, meaning it must be controlled. It has invaded more than 2 million hectares in the country, and its negative impacts include a decrease in native species diversity, an increase in some soil physicochemical properties, and a reduction in grazing area.[5-7] Given the invasion status of *L. camara* in South Africa and the multiple functions of road–river interchanges and roadsides in facilitating alien plant invasion, we aimed to determine if the abundance and invasion of *L. camara* were higher under road–river interchanges than at roadsides. The study is motivated by the need to provide useful road and bridge management information.

Methods and materials

The abundance and invasion extent of *L. camara* along road–river interchanges and roadsides were studied in Soutpansberg and its surroundings in the Vhembe Biosphere Reserve, Limpopo Province of South Africa (Figure 1a). The study area was selected due to the abundance of *L. camara* along roads and the dense national and regional road network. Vegetation in the study area is Soutpansberg Mountain Bushveld in the Soutpansberg mountain area, Makhado Sweet Bushveld and Musina Mopane Bushveld in the south and north of the mountains, respectively.[8] The average annual rainfall is between 340 mm and 1200 mm, being high in the mountainous area and decreasing moving away from the mountainous area. Most rain falls in summer between October and March and temperatures are hot in summer (averaging 40 °C) and mild in winter (averaging 4 °C).[8]

Using roadside surveys, a time and cost-effective method to detect invasive alien plants,[4] a total of 446 km (totalling 48 road–river interchanges) was surveyed along five major roads in the study area. The surveyed roads were: (1) National Road 1 (N1) from Capricorn Tollgate to Musina (143 km), (2) Regional Road 524 (R524) from Makhado (previously known as Louis Trichardt) to Thohoyandou (70 km), (3) Regional Road 523 (R523) from Thohoyandou to Vivo (140 km), (4) Regional Road 578 (R578) from Makhado to Elim (23 km), and (5) Regional Road 522 (R522) from Vivo to Makhado (70 km). Surveys were conducted in winter in July and August 2018 when most road–river interchanges were dry. At each road–river interchange, four plots (two on both sides of the bridge),

each measuring 5 x 5 m, were set up on the riparian zone (approximately 5 m from the bridge wall). These four plots were referred to as road–river interchanges plots. An additional four plots, with similar dimensions, were set up 100 m from the bridge along the road (approximately 5 m from the road) and these were referred to as roadside plots (Figure 1b).

On every road–river interchange where *L. camara* was present, a detailed vegetation survey was conducted in all road–river interchange and roadside plots, resulting in 168 plots being surveyed (8 plots x 21 road–river interchange). Within each plot, species richness and density of all the trees and shrubs were determined from counts of individual plants per plot. In addition, height of all identified individual plants was measured using a ranging pole with 10-cm graduations. Diameter at base for each plant was measured using a Vernier caliper (for multi-stemmed plants only the largest stem was measured). Plant cover was visually estimated as a percentage per plot. All trees and shrubs were collected for identification at the University of Venda herbarium in the Department of Botany.

A *t*-test was performed in Statistica version 13.2 to assess differences between road locations on measured variables. Species richness, Shannon–Wiener diversity index (H') and evenness index (J) were calculated per plot and used to examine species diversity between road location. Data were tested for normality and proof of homogeneity of variances using Kolmogorov–Smirnov and Levene tests, respectively, and data were normally distributed. Species occupancy frequencies (expressed as a percentage) were calculated as the number of times a plant species is present in the different plots. Using the presence and absence data, principal component analysis (PCA) was performed to investigate how plots and species change species composition. PCA was performed using Canoco for Windows version 5.

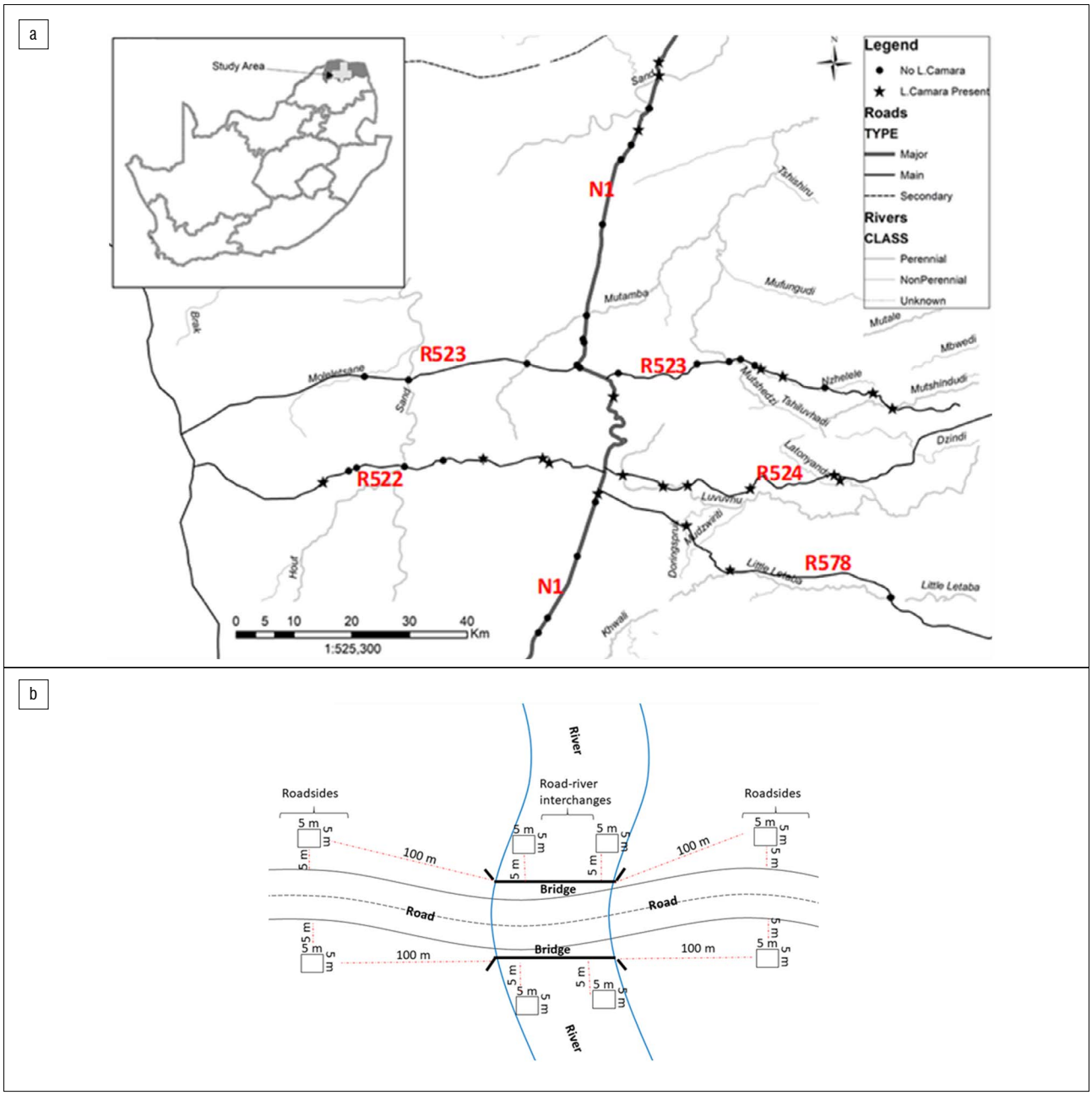


Figure 1: (a) Location of the study area in Vhembe Biosphere Reserve, Limpopo Province of South Africa with surveyed national and regional roads shown. (b) Schematic illustration showing how plots were sampled at different road locations.

Results

Of the 48 road–river interchanges surveyed, 21 (44%) had *L. camara* along both road–river interchanges and roadsides. A total of 1118 *L. camara* individuals were recorded, of which 352 were along road–river interchanges and 766 were along roadsides. Average *L. camara* abundance per plot showed no significant ($t=0.355$; $p=0.725$) differences between road–river interchanges and roadsides (Figure 2a). In contrast, the abundance of other trees and shrubs was significantly ($t=2.588$; $p=0.014$) higher along roadsides than road–river interchanges (Figure 2a). Height of *L. camara* was significantly ($t=1.906$; $p=0.05$) higher along road–river interchanges than roadsides, whereas height of other trees and shrubs was significantly ($t=2.872$; $p=0.006$) higher along roadsides than road–river interchanges (Figure 2b). Diameter at base of *L. camara* was significantly ($t=4.920$; $p=0.001$) higher along road–river interchanges than roadsides. However, diameter at base of other trees and shrubs showed no significant ($t=0.065$; $p=0.851$) differences between road–river interchanges and roadsides (Figure 2c). Cover of both *L. camara* and other trees and shrubs showed no significant ($p>0.05$) differences between road–river interchanges and roadsides (Figure 2d).

Both species richness and Shannon–Wiener showed significant ($p=0.001$) differences between road–river interchanges and roadsides (Figures 3a and b). The above-mentioned differences in species richness and Shannon–Wiener showed greater diversity along roadsides than road–river interchanges (Figure 3a and 3b). However, there were no significant ($p=0.582$) differences in species evenness between road–river interchanges and roadsides (Figure 3c). A total of 29 trees and shrub species were recorded in all plots (Supplementary table 1). Besides *L. camara* dominating road–river interchanges, other species with more than 60% occupancy frequency were the alien species *Solanum mauritianum, Caesalpinia decapetala* and *Rubus rigidus* and the native species *Diospyros lycioides*, *Euclea natalensis*, *Lippia javanica*, *Podocarpus* sp. and *Vachellia karroo* (Supplementary table 1). Similarly, *L. camara* dominated roadsides, and other species with occupancy frequency greater than 80% were the alien species *Acacia mearnsii* and native species *D. lycioides*, *E. natalensis*, *L. javanica*, *Podocarpus* sp., *Searsia pentheri* and *V. karroo* (Supplementary material 1). Five native species, namely *Combretum vendae*, *Englerophytum magalismontanum*, *E. divinorum*, *P. latifolius* and *Zanthoxylum capense* were not present along road–river interchanges but appeared along roadsides with an occupancy frequency of less than 60%. PCA showed little separated road–river interchanges and roadsides in ordination space determined by the first two axes between them (Figure 3d). Nevertheless, clear distinctions can be seen regarding assemblages for some species, for example, *S. pentheri*, *P. latifolius* and *Psidium guajava* assembled more along roadsides than did *S. mauritianum,* which assembled more along road–river interchanges (Figure 3d).

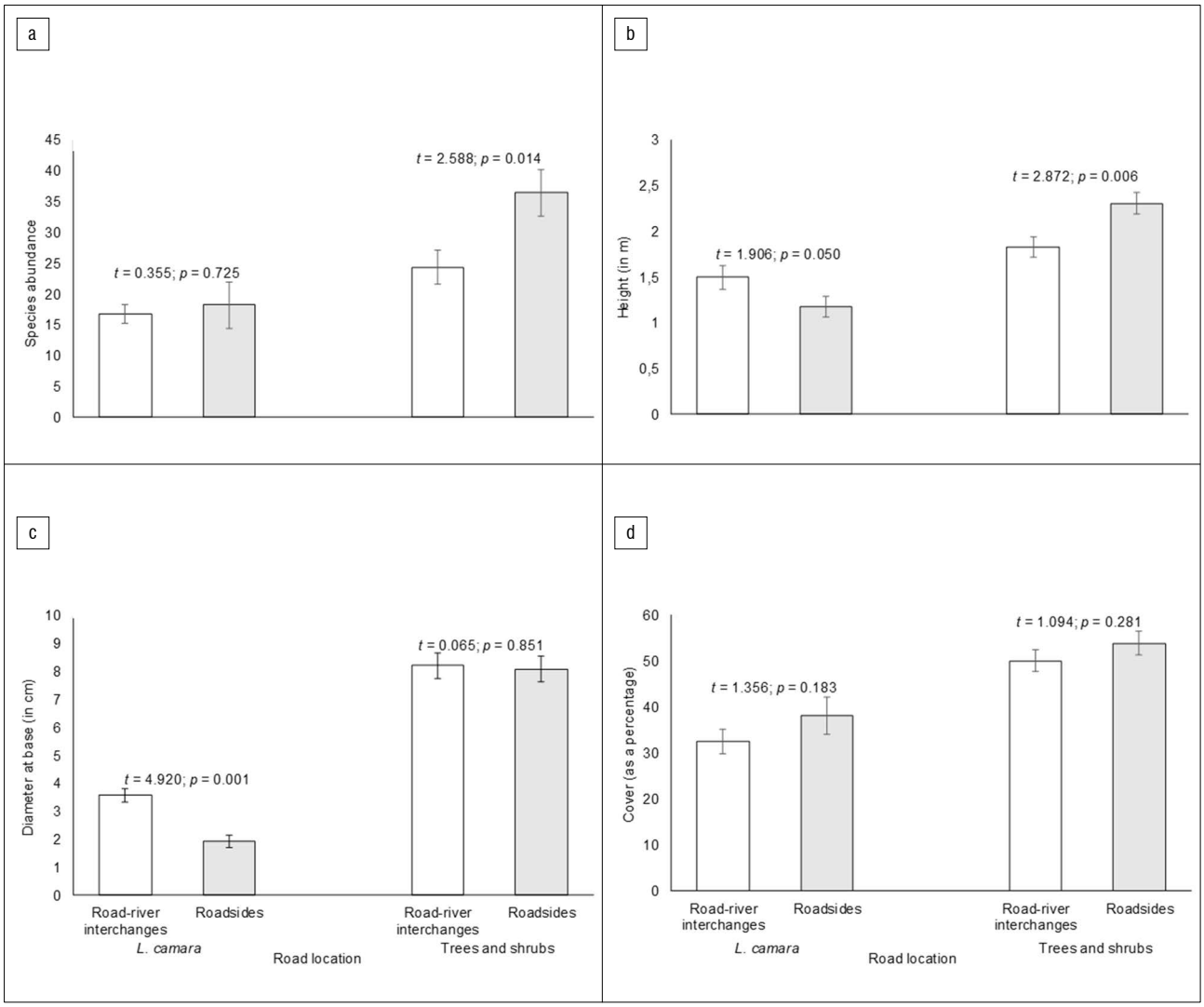


Figure 2: Comparison of (a) species abundance, (b) height, (c) diameter at base and (d) cover for *L. camara* and other trees and shrubs at different road locations. Data are means ± s.e. and *t*-test results are shown.

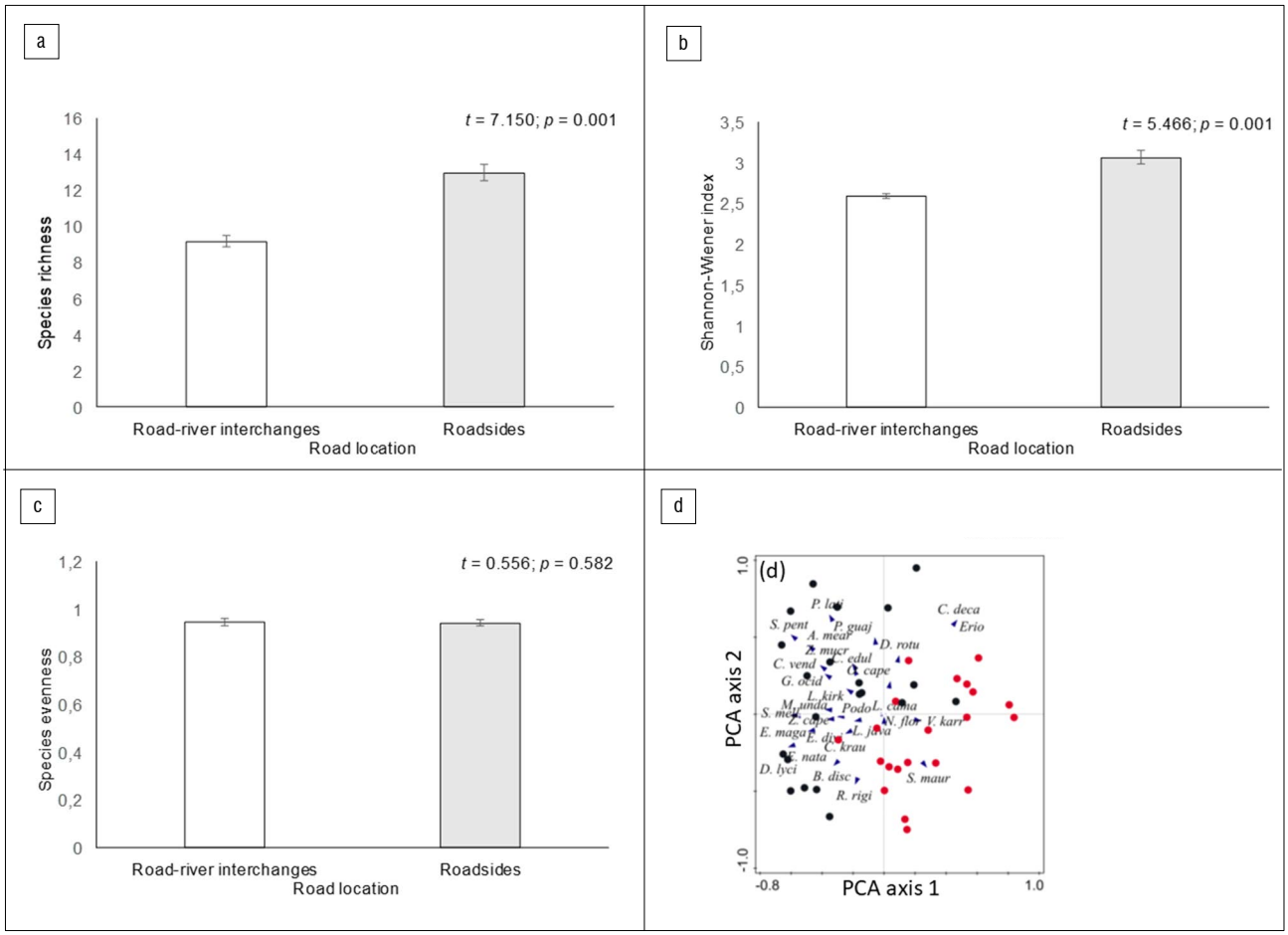


Figure 3: Comparison of (a) species richness, (b) Shannon–Wiener index and (c) evenness at different road locations. Data are means ± s.e. and *t*-test results are shown. (d) Principal component analysis (PCA) of species (▲) and road location as (●) road–river interchanges and (●) roadsides for identified trees and shrubs. The first four letters of the species names are presented, with full names given in Supplementary table 1.

Discussion

In this study, *L. camara* abundance and cover were similar between road–river interchanges and roadsides. Height and diameter of *L. camara* were greater along road–river interchanges than roadsides – an indication that the species grows better at road–river interchanges than roadsides. Therefore, it is evident that road–river interchanges facilitate *L. camara* invasion and create suitable conditions for the species to establish. This result concurs with previous studies on other invasive species, e.g. *Pennisetum setaceum* which dominates South African roadsides but grows better under road–river interchanges.[3] Similarly, Baard and Kraaij[4] identified 109 invasive plant species along the Garden Route roads, with the highest incidences being on degraded and transformed lands as well as farm roads. Other studies also reported that invasive plants such as *Clidemia hirta* dominate roads and trails in Malaysia[9], whilst *Ambrosia artemisiifolia* dominate paved roads in Canada[10]. We suspect that both road–river interchanges and roadsides contribute to invasion by invasive alien plants through (1) provision of migration corridors, (2) creation of microcosm environments (especially road–river interchanges) that are suitable for invasive plant growth and establishment, and (3) road construction and maintenance disturbance which create conditions for alien plant establishment through soil movement.[1,3,11,12] In addition, several environmental factors (e.g. soil type, flooding and climate change) can influence the abundance and spread of invasive plants along road–river interchanges and roadsides.[3] Given the fact that other invasive plants like *S. mauritianum, C. decapetala, A. mearnsii* and *R. rigidus* were present at surveyed road locations, our observations are consistent with the suggestion that road–river interchanges and roadsides facilitate invasion by invasive alien plants.[3,12] Baard and Kraaij[4] identified 35 sleeper weeds that displayed invasion plant tendencies but are not listed by the South African Plant Invader Atlas or regulated by legislation. In this study, the occurrence of other woody invaders on surveyed plots seems to suggest that road–river interchanges and roadsides, like any other disturbance mechanism, promotes co-occurrence of invaders.

Although roadside surveys have disadvantages – e.g. poor detection of plants far from the road verge and a bias towards large growth form – results of this study suggest that the abundance and invasion of *L. camara* is concentrated along both road–river interchanges and roadsides, an indication that roads and bridges play a key role in facilitating *L. camara* invasion and establishment. The presence of *L. camara* along road–river interchanges and roadsides poses a major problem as the plant can produce more propagules that might be dispersed along the road and river, resulting in *L. camara* expansion. Because roads and bridges are important within the national infrastructure, proper management of *L. camara* and other invasives along roads and bridges is required. We recommend that road–river interchanges and roadsides be prioritised during invasive alien plant control in South Africa.

Acknowledgements

We thank the DST-NRF Centre of Excellence for Invasion Biology and the National Research Foundation of South Africa (Thuthuka grant TTK190117408656) for funding to conduct field work.

Competing interests

We declare that there are no competing interests.

Authors' contributions

E.S.M.: Collected the data and drafted the manuscript. S.R.: Conceptualised the study, methodology, funding acquisition, verified the data set used in the final analysis, participated in the revision of the manuscript, student supervision, and project leadership and management.

References

1. Greenberg CH, Crownover SH, Gordon DR. Roadside soils: A corridor for invasion of xeric scrub by nonindigenous plants. Nat Areas J. 1997;17:99–109.
2. Jodoin Y, Lavoie C, Villeneuve P, Theriault M, Beaulieu J, Belzile F. Highways as corridors and habitats for the invasive common reed *Phragmites australis* in Quebec, Canada. J Appl Ecol. 2008;45:459–466. https://doi.org/10.1111/j.1365-2664.2007.01362.x
3. Rahlao SJ, Milton SJ, Esler KJ, Barnard P. The distribution of invasive *Pennisetum setaceum* along roadsides in western South Africa: The role of corridor interchanges. Weed Res. 2010;50:537–543. https://doi.org/10.1111/j.1365-3180.2010.00801.x
4. Baard JA, Kraaij T. Use of a rapid roadside survey to detect potentially invasive plant species along the Garden Route, South Africa. Koedoe. 2019;61(1), a1515. https://doi.org/10.4102/koedoe.v61i1.1515
5. Vardien W, Richardson DM, Foxcroft LC, Thompson GD, Wilson JRU, Le Roux JJ. Invasion dynamics of *Lantana camara* L. (sensu lato) in South Africa. S Afr J Bot. 2012;81:81–94. https://doi.org/10.1016/j.sajb.2012.06.002
6. Van Wilgen BW, Reyers B, Le Maitre DC, Richardson DM, Schonegevela L. A biome scale assessment of the impact of invasive alien plants on ecosystem services in South Africa. J Environ Manage. 2008;89:336–349. https://doi.org/10.1016/j.jenvman.2007.06.015
7. Ruwanza S, Shackleton CM. Effects of the invasive shrub, *Lantana camara*, on soil properties in the Eastern Cape, South Africa. Weed Biol Manag. 2016;16:67–79. https://doi.org/10.1111/wbm.12094
8. Mucina L, Rutherford MC. The vegetation of South Africa, Lesotho and Swaziland. Strelitzia 19. Pretoria: South African National Biodiversity Institute; 2006.
9. Le C, Fukumori K, Hosaka T, Numata S, Hashim M, Kosaki T. The distribution of an invasive species, *Clidemia hirta* along roads and trails in Endau Rompin National Park, Malaysia. Trop Conserv Sci. 2018;11:1–9. https://doi.org/10.1177/1940082917752818
10. Joly M, Bertrand P, Gbangou RY, White M-C, Dube' J, Lavoie C. Paving the way for invasive species: Road type and the spread of common ragweed (*Ambrosia artemisiifolia*). Environ Manage. 2011;48:514–522. https://doi.org/10.1007/s00267-011-9711-7
11. Gelbard JL, Belnap J. Roads as conduits for exotic plant invasions in a semiarid landscape. Conserv Biol. 2003;17:420–432. https://doi.org/10.1046/j.1523-1739.2003.01408.x
12. Von der Lippe M, Kowarik I. Long-distance dispersal of plants by vehicles as a driver of plant invasions. Conserv Biol. 2007;21:986–996. https://doi.org/10.1111/j.1523-1739.2007.00722.x