South African Journal *of* Science

volume 116
number 3/4

Influence of race and collaboration on research publication

Gender pay transparency mechanisms for South Africa

A lens on materials science

Significance of the coelacanth sighting on the South Coast

A disciplinary discussion on decolonisation

ASSAf
ACADEMY OF SCIENCE OF SOUTH AFRICA

Volume 116
Number 3/4
March/April 2020

South African Journal of Science

eISSN: 1996-7489

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

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
Coelacanth sighted off Pumula on the KwaZulu-Natal South Coast, South Africa, on 22 November 2019 (photo: Bruce Henderson). In a Commentary on page 19, Bruton et al. describe the sighting and its significance.

Sustainability in an era of emerging infectious diseases

The outbreak of the coronavirus COVID-19 earlier this year took the world by surprise.The long-term consequences of this emerging infectious disease (EID) are impossible to predict. Currently, disruption in every economic sector, unprecedented demands on health services everywhere, let alone human morbidity and mortality, are already highly visible through saturated media coverage. The Academy of Science of South Africa recently issued a measured Statement on the Implications of Novel Coronavirus (SARS-CoV-2; COVID-19) in South Africa that is worth reading. A recent 'Opinion' in the *Proceedings of the National Academy of Sciences of the United States of America* alerts us to the effects of a pandemic risk that threatens progress towards all the Sustainable Development Goals (SDGs).[1]

A number of articles in this issue of SAJS are of relevance to the topic of SDGs. One that now takes on a fresh resonance in the age of EIDs is the review by Catherine Burns of Howard Phillips's book *In a Time of Plague: Memories of the 'Spanish' Flu Epidemic of 1918 in South Africa.* As an historian of medicine, Phillips has contributed to the literature on this flu epidemic[2,3], but here he presents personal, first-hand accounts of the impact of the epidemic in South Africa. Drawing on interviews and letters, and quoting poetry, songs and reminiscences of an unforgettable period of disease that claimed the lives of 50 million people worldwide[4], this book has implications for reaching SDG 3, 'Good health and well-being', with our own imminent threats of large-scale illnesses and limited resources with which to treat them.

Other Book Reviews also relate to SDGs. Hettie Schönfeldt analyses the late Renata Coetzee's *A Feast from Nature: Food Culture of the First Humans on Planet Earth* that aligns with SDG 2, 'Zero hunger'. Should food imports be affected by EIDs, greater knowledge, and more appreciation, of local crop varieties and indigenous animal breeds gains increased significance. SDG 4, 'Quality education', is the topic of Chris Brink's *The Soul of a University: Why Excellence Is Not Enough*, reviewed by Harry Boyte. Already challenged by student disruptions, it is certain that EIDs will create even more difficulties to the education sector. Sydney Moyo reviews an important reference work relating to SDG 6, 'Clean water and sanitation'. *Multiple Stressors in River Ecosystems: Status, Impacts and Prospects for the Future*, edited by Sergi Sabater, Arturo Elosegi and Ralf Ludwig, includes the latest knowledge of the pressures facing the ecology and management of running waters, while Alan Whitfield's *Fishes of Southern African Estuaries: From Species to Systems*, reviewed by Peter Moyle, alerts us to the importance of SDG 14, 'Life below water'. There are also three Commentaries that highlight other SDGs. One, 'All materials great and small', is by Alexander Quandt of the School of Physics at Wits and the 2018/2019 NSTF-South32 Special Annual Theme Award winner. In charting his own career in the field of materials science, Quandt highlights the critical importance of innovation and creative science that relates to SDG 9, 'Industry, innovation and infrastructure. Nithaya Chetty, Dean of Science at the University of the Witwatersrand, also gives pause for thought to SDG 9, from the perspective of the university curriculum and his essay also has implications for SDG 4, 'Quality education'. In 'Re-inventing inventiveness in science', Chetty argues that the future of the South African academy and economy depends on training an increasing number of highly qualified postgraduates who can readily be employed outside academia and be sought after for their innovative perspectives. The other, titled 'From reindeer to rhino: Reflections on 'Climate change mitigation and adaptation benefits of wilder rangelands'', is by Graham Kerley and Joris Cromsigt who initiated a network of scientists and managers in South Africa and Sweden to refresh the rangeland debate in South Africa in the face of climate change. Theirs is a contribution to SDG 13, 'Climate action', as well as to SDG 15, 'Life on land'.

Research Articles in this issue also speak powerfully to the SDGs. SDG 16, 'Peace, justice and strong institutions', is the theme of Sikanyiso Masuku and Sharmla Rama's contribution, 'Government and civil society in South Africa: Collaboration and challenges in securing refugee rights'. They present a case study of the numerous bureaucratic and institutional difficulties that confront Congolese refugees in Pietermaritzburg in securing their legal rights to employment, education and health care. 'Gender pay transparency mechanisms: Future directions for South Africa', by Anita Bosch and Shimon Barit, directly addresses SDG 5, 'Gender equality'. As is well known, gender pay discrimination remains a stumbling block in achieving gender equality and by making gender wage differentials more transparent, employers may be compelled to remunerate fairly and equally.

Matthew McGill and co-authors contribute 'Observation and quantification of aerosol outflow from southern Africa using spaceborne lidar', relating to SDG 7, 'Affordable and clean energy'. They demonstrate how biomass burning in Africa provides a prolific source of aerosols that are transported from source to very distant areas. Models have long predicted the primary outflow and transport routes, but lidar on the International Space Station has detection sensitivity that supports long-held theories of aerosol transport from the African subcontinent as far downstream as Australia. 'Sciarid pests (Diptera: Sciaridae) from undercover crop production in South Africa' by Agil Katumanyane and co-authors homes in on SDG 2, 'Zero hunger'. In this multidisciplinary contribution the authors explain how and why fungus gnats (sciarids) are some of the most important pests in undercover crop production, causing direct physical damage to plant roots, transferring fungal pathogens and creating entry points for soil-borne plant pathogens.

The research article by Omosalewa Odebiri and co-authors has implications for SDG 13, 'Climate action' and also for SDG 15, 'Life on land'. Titled 'Estimating soil organic carbon stocks under commercial forestry using topo-climate variables in KwaZulu-Natal, South Africa', they emphasise soil as the largest carbon reservoir but point out that the relationship between soil organic carbon in commercial forests and topo-climatic variables is still poorly understood. They argue that bridging this knowledge gap is essential for quantifying local, regional or global carbon balances.

The coronavirus outbreak has not only reminded us that humanity remains in what John Kenneth Galbraith described in 1977 as *The Age of Uncertainty*[5], and that our socio-economic structures are fragile, but also that there is even more urgency to make progress in reaching sustainable development.

References

1. Di Marco H, Baker ML, Daszak P, De Barro P, Eskew EA, Godde CM, et al. Opinion: Sustainable development must account for pandemic risk. Proc Natl Acad Sci USA. 2020;117(8):3888–3892. https://doi.org/10.1073/pnas.2001655117
2. Phillips H, Killingray D, editors. The Spanish influenza pandemic of 1918–19: New perspectives. London: Routledge; 2003.
3. Phillips H. Black October: The impact of the Spanish influenza epidemic of 1918 on South Africa. Pretoria: Government Printer; 1990.
4. US Centers for Disease Control and Prevention. 1918 Pandemic (H1N1 virus) [webpage on the Internet]. No date [updated 2019 Mar 20; cited 2020 Mar 03]. Available from: https://www.cdc.gov/flu/pandemic-resources/1918-pandemic-h1n1.html
5. Galbraith JK. The age of uncertainty. Boston: Houghton Mifflin; 1977.

Note: Issue 116(1/2) included an Invited Commentary by Fortunate M. Phaka on 'Environmental science investigations of folk taxonomy and other forms of indigenous knowledge'. We have been asked to note formally that this contribution was not peer reviewed and that the images used were borrowed from an as-yet unpublished paper by Phaka.

HOW TO CITE:
Carruthers J. Sustainability in an era of emerging infectious diseases. S Afr J Sci. 2020;116(3/4), Art. #8043, 1 page. https://doi.org/10.17159/sajs.2020/8043

AUTHORS:
Dirk Roux[1,2]
Hayley Clements[3]
Bianca Currie[2]
Hervé Fritz[2,4]
Paddy Gordon[5]
Nerina Kruger[6]
Stefanie Freitag[5]

AFFILIATIONS:
[1]Scientific Services, South African National Parks, George, South Africa
[2]Sustainability Research Unit, Nelson Mandela University, George, South Africa
[3]Centre for Complex Systems in Transition, Stellenbosch University, Stellenbosch, South Africa
[4]REHABS International Research Laboratory, Nelson Mandela University, George, South Africa
[5]South African National Parks, Knysna, South Africa
[6]Scientific Services, South African National Parks, Sedgefield, South Africa

CORRESPONDENCE TO:
Dirk Roux

EMAIL:
Dirk.Roux@sanparks.org

HOW TO CITE:
Roux D, Clements H, Currie B, Fritz H, Gordon P, Kruger N, et al. The GRIN Meeting: A 'third place' for managers and scholars of social-ecological systems. S Afr J Sci. 2020;116(3/4), Art. #7598, 2 pages. https://doi.org/10.17159/sajs.2020/7598

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
sustainability science, science–management interface, co-learning, postgraduate Spring School

PUBLISHED:
26 March 2020

The GRIN Meeting: A 'third place' for managers and scholars of social-ecological systems

In his book *The Great Good Place*[1], Ray Oldenburg writes about the important role that 'third places' play in community building. Third places refer to social environments other than home (first place) and work (second place) where people spend time. These places provide neutral ground for engagement and relationship building, with conversation or dialogue being the main activity.[1] Churches, cafes, clubs, public libraries and parks are examples of third places. The ancient Greek concept of the *agora* or 'gathering place' is an older variant of a third place, which served as a dedicated public space in a city-state for deliberating philosophical, artistic, spiritual, economic and political affairs of the day.

Third places or *agora*-type spaces are also required for scientists and practitioners to meet, share experiences and learn together.[2] This requirement is particularly relevant for addressing sustainability challenges, which typically requires consideration of diverse scientific expertise in combination with social values, policies and management practices.[3] Sustainability challenges are social-ecological in nature and characterised by complex feedbacks, contested decision options and uncertain outcomes. Addressing these challenges usually goes beyond the capacity of any one organisation[4] and requires integration of knowledge from across disciplines, sectors, scales and science–policy–practice realms.[5] Furthermore, these challenges tend to evolve and hence call for adaptive approaches that allow for continual reframing, ongoing and collaborative learning and cooperative decision-making.[6]

The Garden Route Interface and Networking (GRIN) Meeting was initiated in 2017 with the aim of creating a third place for dialogue on, and advancement of, research and practice dealing with the interactions between natural and social systems, and with how those interactions affect the challenge of sustainability. Held in the Garden Route over 3 days during September–October of each year, GRIN addresses two important interfaces: research–practice and human–nature (or social-ecological). As such, the overarching theme of GRIN meetings is *research and management for sustainable social-ecological systems.* Social-ecological systems that have been presented at GRIN meetings include estuaries, protected areas, agricultural landscapes, cities, fisheries and catchments.

During its first 2 years, the Meeting was referred to as the Garden Route Interface Meeting (GRIM), but delegates voted for a happier acronym (GRIN) during the 2019 gathering. To date, GRIN has attracted about 70 delegates per year, including from South Africa (>50%), France, Finland, Spain, Canada, Kenya and Nigeria. GRIN is jointly organised and hosted by South African National Parks (SANParks), Nelson Mandela University (specifically its Sustainability Research Unit located at the George Campus), the French National Centre for Scientific Research (specifically their International Research Laboratory REHABS located at the George Campus of Nelson Mandela University, with financial support from the French Embassy) and the Southern African Programme on Ecosystem Change and Society.

Feedback from previous meetings highlight that the most valued features of GRIN are: (1) active engagement of managers/practitioners (>20% of delegates) through discussions and presentations; (2) ample time for networking and socialising, such as that during a GRIN braai, a 5-km sunset walk or a mid-meeting field excursion; and (3) presentations on the latest developments in social-ecological systems thinking. Presentation themes that have emerged to date and that remain relevant for future meetings are: understanding of human–nature connections (including through the use of social media data); strategies used to promote knowledge exchange among science, policy and practice; sustainable resource use and balancing benefits and conflicts related to services from nature; multiscale governance challenges (from rural towns to estuaries and peace parks); establishing long-term social-ecological research sites; resilience, adaptations and transformations in social-ecological systems; responding to climate and other drivers of change; social networks, social learning and collective action; and methods/approaches for studying social-ecological systems.

Because 'sustainability science' is a young and emerging field of study that departs substantially from conventional and disciplinary modes of inquiry (see Haider et al.[7]), the increasing number of postgraduate students pursuing studies in this field may not find optimal exposure in their often disciplinary (whether from social or natural sciences) university departments. GRIN provides a space for mutual learning about experiences such as applying mixed-methods research (combining qualitative and quantitative techniques), engaging stakeholders and facilitating knowledge co-production, and the challenges of doing a transdisciplinary PhD. To further cater for the specific needs that are commonly expressed by these students, a multi-day Spring School (2018) and half-day workshop (2019) on social-ecological systems research have been presented back-to-back with previous GRIN meetings.

GRIN 2020 is scheduled to take place from 6 to 8 October at Pine Lake Marina near Sedgefield (in the Western Cape, South Africa), and is expected to be followed by a Spring School on social-ecological systems research. A formal announcement and first call for abstracts will be circulated by the end of the first quarter of 2020. Detailed information about the 2020 meeting and Spring School can be found at GRIN2020.

References

1. Oldenburg R. The great good place: Cafes, coffee shops, community centers, beauty parlors, general stores, bars, hangouts, and how they get you through the day. New York: Paragon House; 1989. https://doi.org/10.2307/2072095
2. Roux DJ, Nel JL, Cundill G, O'Farrell P, Fabricius C. Transdisciplinary research for systemic change: Who to learn with, what to learn about and how to learn. Sustain Sci. 2017;12(5):711–726. https://doi.org/10.1007/s11625-017-0446-0

3. Wyborn C, Datta A, Montana J, Ryan M, Leith P, Chaffin B, et al. Co-producing sustainability: Reordering the governance of science, policy, and practice. Annu Rev Env Resour. 2019;44:319–346. https://doi.org/10.1146/annurev-environ-101718-033103

4. Game ET, Meijaard E, Sheil D, McDonald-Madden E. Conservation in a wicked complex world: Challenges and solutions. Conserv Lett. 2014;7(3):271–277. https://doi.org/10.1111/conl.12050

5. Van Kerkhoff L. Developing integrative research for sustainability science through a complexity principles-based approach. Sustain Sci. 2014;9(2):143–155. https://doi.org/10.1007/s11625-013-0203-y

6. Sayer J, Sunderland T, Ghazoul J, Pfund JL, Sheil D, Meijaard E, et al. Ten principles for a landscape approach to reconciling agriculture, conservation, and other competing land uses. Proc Natl Acad Sci USA. 2013;110:8349–8356. https://doi.org/10.1073/pnas.1210595110

7. Haider LJ, Hentati-Sundberg J, Giusti M, Goodness J, Hamann M, Masterson VA, et al. The undisciplinary journey: Early-career perspectives in sustainability science. Sustain Sci. 2017;13(1):191–204. https://doi.org/10.1007/s11625-017-0445-1

AUTHOR:
Wouter Holleman[1]

AFFILIATION:
[1]South African Institute for Aquatic Biodiversity, Makhanda, South Africa

CORRESPONDENCE TO:
Wouter Holleman

EMAIL:
w.holleman@saiab.ac.za

HOW TO CITE:
Holleman W. Phillip C. Heemstra (1941–2019): Ichthyologist extraordinaire. S Afr J Sci. 2020;116(3/4), Art. #7812, 2 pages. https://doi.org/10.17159/sajs.2020/7812

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

PUBLISHED:
26 March 2020

Phillip C. Heemstra (1941–2019): Ichthyologist extraordinaire

When Margaret Smith was appointed the first director of the JLB Smith Institute of Ichthyology (now the South African Institute for Aquatic Biodiversity) in Makhanda (formerly Grahamstown) in 1968 she was determined to build on the legacy of her husband, Professor JLB Smith. In 1978, she initiated a complete revision of his famous book, *Sea Fishes of Southern Africa*, initially planning to handle the whole revision herself with a research assistant. She soon realised that it would be impossible to do so and so, at the recommendation of Dr John Randall of Hawaii, she appointed Dr Phil Heemstra, a highly respected taxonomist from the University of Florida in the USA as her co-editor. This was a master stroke as Heemstra – approachable, meticulous, dedicated, talented – proved to be a huge asset.

Taking into account the specialisation that had taken place in ichthyology in recent years, Smith and Heemstra soon realised that they could not produce *Sea Fishes* alone, so they enlisted the services of the leading international authorities on each fish family occurring in the southern African region. The book eventually involved 77 collaborators from 15 countries and was titled *Smiths' Sea Fishes* to acknowledge the contributions of both Smiths. The multi-authored volume covered 2150 fish species in 270 families, with Heemstra authoring or co-authoring an impressive 74 family accounts. In 1990, Heemstra made a further important contribution to marine fish taxonomy by co-editing, with Ofer Gon, *Fishes of the Southern Ocean*, in which he authored 7 of the 49 family accounts. His legacy will be confirmed with the publication of *Coastal Fishes of the Western Indian Ocean* later in the year, successor to *Smiths' Sea Fishes*, and for which he authored more than 70 of the 260 family accounts.

At the time of Heemstra's formal retirement in March 2002, several colleagues published tributes to him in the Ichthyology Institute's newsletter, *Ichthos*, that emphasise his mentorship and guidance:

> *I was privileged to have had Phil Heemstra as the Senior Marine Scientist throughout my term as Director. Phil was the unquestionable successor to the legendary JLB Smith, and shared many of JLB's characteristics – tenacity, utter dedication, a deep love of his work, and an obstinate determination to do what had to be done no matter how many obstacles were placed in his path. Pinned on to his wall was a saying that epitomised his approach to life: 'Yard by yard, life is hard; inch by inch, it's a cinch'. I believe that a good taxonomist – as was Phil – needs to have particular traits in order to succeed: meticulous to a fault, an ability to concentrate on minutiae while always keeping the big picture in mind, and a thorough understanding of the theory behind the practice.*
>
> Professor Mike Bruton, Director of the Ichthyology Institute, 1982–1994

> *Phil's greatest accomplishment was his collaboration with Margaret Smith and many colleagues around the world to produce a fresh version of* Smiths' Sea Fishes. *This monumental work, widely proclaimed as 'the fisherman's bible' and 'the best book of its kind in the world', involved not only world-class research but also a truly spectacular level of project management and coordination. He also produced a steady stream of original publications on taxonomy and systematics in peer-reviewed journals and was the expert on fishes of the Western Indian Ocean.*
>
> *Not content with merely again revising* Smiths' Sea Fishes *in 1996 Phil embarked on another major challenge – to document the fishes of the Western Indian Ocean (WIO). In doing so he followed a path well worn by the* Smiths' Sea Fishes *project, by inviting recognized world authorities to revise the families in the WIO and prepare accounts as chapters for the proposed book. However, Phil wrote most of the chapters himself.*
>
> *Other major achievements included co-editing the authoritative* Fishes of the Southern Ocean *(with Ofer Gon),* Groupers of the World *(with Matt Craig and Yvonne Sadovy), and numerous scientific papers. He also served as editor of the Ichthyology Institute's scientific publications.*
>
> *Heemstra participated in many expeditions to various sites in the Western Indian Ocean (WIO), from the Red Sea to Mozambique, the Comoros Islands, Madagascar, Mauritius and Rodrigues. These and other endeavours ensured that he became the world's leading authority on the fishes of the WIO. Together with his wife, Elaine, he developed 'Fishwatch', a pioneering science communication project through which recreational SCUBA divers were encouraged to identify and record fishes observed on their dives… 'Fishwatch' expeditions became an important source of original new material in the form of specimens and photographs for the revision of the fishes of the WIO.*
>
> *When Phil formally retired, in order that his wealth of knowledge would not be lost, he was created the Institute's first Emeritus Marine Scientist, allowing him to focus on the WIO book. Under Phil's fastidious editorship the Institute's publications enjoyed a fine international reputation and a high standard. These and other strengths are ultimately what made him a fine marine systematist.*
>
> Professor Paul Skelton, Director/Managing Director of the Institute, 1995–2011

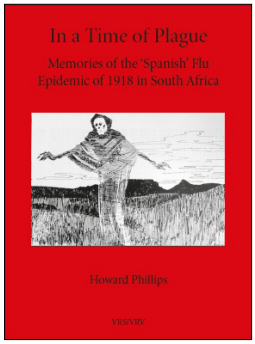
I worked as Phil Heemstra's mangrove mud Research Assistant for 21 years. I was initially employed by Margaret Smith while Phil was on sabbatical leave in the USA and Phil was merely told that he had an assistant to help with revising Sea Fishes. *Phil taught me everything I know about fish identification and all other aspects relating to fish taxonomy, methods of cleaning and staining, skeletal preparations, etc., and I could not have had a more patient and thorough teacher. He never wearied of answering the same questions repeatedly and was always willing to peer down a microscope and check my identifications or correct my mistakes. In the beginning, while I was still helping Margaret Smith, I would approach him and announce that I could not see the lateral line pores or count the gill rakers, or something like that. He always replied, 'Don't you know that she only gives you the difficult ones she can't do herself?' Being involved in the various stages of the production of* Smiths' Sea Fishes *was a tremendous learning experience, and we all felt a great sense of achievement when it was finally published … I remember someone asking him many years ago whether he ever tired of fishes. His reply described him in a nutshell: 'I am eternally fascinated by fishes.'*

Joan Wright, Heemstra's Research Assistant

One day, asking him if he would be able to complete a particular section for the WIO fishes book, his answer was, 'It's do-able'. And that was Phil – you got on with the job at hand and DID IT. He epitomised the words of Elizabeth Barrett Browning: 'Great things must be done greatly, with a great purpose, a great mind, a great courage, a great energy and a great persistent patience.'

Wouter Holleman, Research Associate, South African Institute for Aquatic Biodiversity

BOOK TITLE:
In a time of plague: Memories of the 'Spanish' flu epidemic of 1918 in South Africa



EDITOR:
Howard Phillips

ISBN:
9780994720719 (hardcover, 189 pp)

PUBLISHER:
Van Riebeeck Society, Cape Town; ZAR380

PUBLISHED:
2018

REVIEWER:
Catherine Burns

AFFILIATION:
Adler Museum of Medicine, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa

EMAIL:
catherine.burns@wits.ac.za

HOW TO CITE:
Burns C. 'It chose the beautiful ones….'. S Afr J Sci. 2020;116(3/4), Art. #7717, 2 pages. https://doi.org/10.17159/sajs.2020/7717

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

PUBLISHED:
26 March 2020

'It chose the beautiful ones….'

> *It was only the illness and dying of people that I remember. And the funerals, Oh! My brother Elias* [Plaatje, an evangelist] *had just caught the 'flu and died in the same 'flu. The 'flu really comes with the soldiers from overseas. Because people didn't even know what was causing that fever. Oh, it was terrible. People working in the mines who had died – they just made one grave for them. The children eight to twenty were mostly dying. Even the doctors could do nothing... I was supposed to get married in 1918, then the 'flu came and I got sick. I was in Kimberley for buying my wedding gown... the 'flu was very strong over September and October and over by November and towards December. ...Even the Europeans in Barkly West had to collect this 'wel-als' bush. You only collect that bush in the veld and boil it...*
>
> Martha Bokako (born 1890), interviewed 22 November 1980 by Andrew Reed (p.95)

Martha Bokako was mourning her family's deep loss, and recovering from a serious bout of 'flu in Pniel (Western Cape) in late 1918, and yet she recalled anticipating the life ahead of her – including her hope that her postponed wedding plans would resume when her fiancée, an ordained minister, had recovered from the most thoroughgoing and severe epidemic the region had ever faced. In the 3 months Martha Bokako refers to here, South Africa lost about 4% of the total estimated population; millions of loved ones were left in mourning and many children were left orphaned and unmoored. Her full account, with that of 127 of her fellow South Africans, has been collected by Professor Howard Phillips of the Department of History at the University of Cape Town in *In a Time of Plague: Memories of the 'Spanish 'Flu Epidemic of 1918 in South Africa*. This unique, harrowing and deeply engaging collection, has been edited and brought to life with a lucid and crisp introduction, carefully annotated throughout, to bring the specific geographical, medical and social details of the hundreds of witnesses to this epidemic into our world a century later. The letters and accounts in this collection locate South Africa in a tri-continental frontier – a region deeply imbricated in global movements of people, goods, animals, ideas, ideologies, forms of exchange and extraction. With people and microbes came also their microscopic parasitic cousins – viruses. The Union of South Africa, established in 1910, was one of the newest political formations in the world and the meeting place for people of the sub-continent, from Europe, and, with the outbreak of World War I in 1914, with new arrivals and cargoes from Atlantic African ports, from east Africa and the Mediterranean, and from ports and towns along the Indian Ocean. This book traces the routes of the epidemic through the words and memories of witnesses and survivors in accounts gathered in the 1970s and 1980s and is contextualised with painstaking archival research.

This book is the work of a lifetime of reading, collecting and analysing by Howard Phillips. In his earlier seminal scholarly works[1,2] on the influenza epidemic in South Africa, Phillips demonstrated that the race- and class-based official responses were rooted in the approaches of neo-colonialism by health and state planners of the day. He has shown how these were mitigated, but not obviated, by many acts of civic unity and generosity, courage, and shared resources and responses. One consequence of the post-epidemic era was the state, and elite society, doubling down on segregation as a state planning response to disease vulnerability and contagion. Phillips has also shown that the epidemic forced certain contradictions into the forefront of state health planning – initiating a set of universalistic public health goals; crafting policy and legislation around health provision; and acknowledging key individual and population-based health needs. This propelled the first county-wide health architecture, which, over time, led to providing better national care, even in the face of continuing gross iniquitous spending and racially divided facilities and institutions. A glaring example of such a contradiction is that even as this influenza epidemic devastated natal health, and undermined many generations of women's faith in the efficacy of medical institutions, it also initiated the first state (rather than the hitherto missionary institution-dominated) efforts at comprehensive pregnancy, immunisation and infant care.

In this collected volume, ordinary people's interpretations and voices are foregrounded. Martha Bokako's eloquent account of the 'flu epidemic in South Africa that struck in 1918, with the horror of October in her mind, echoes the many traumatic accounts, recorded in handwritten and typed letters and detailed interview-based recollections. People from the region – from the Cape Province through the Karoo across to communities on the Eastern seaboard and up into the interior of the country – including the huge conurbation of the Witwatersrand, responded to queries from researchers in heartfelt texts, or shared their memories in person.

It is rare in South African scholarship to have a wide cross-section of ordinary people's responses to an event or a calamity gathered in one collection. Coming alive in these pages are the causes and ramifications that Phillips summarises in his Introduction, and in the pages that follow, these searing memories form part of a wider whole. We begin to see a citizenry speaking in chords – experiencing and remembering a time when, for all that separated them, the memories and spectral images of this event coalesced and aggregated. Through Xhosa, Tswana, Afrikaans and Zulu into English translations we hear the ring of specific haunting phrases, cast around starkly similar factual accordances and details. In uncannily similar metaphorical phrases we understand that the epidemic was a shared national event, and how people at the time – even as children – understood that it was linked both to the dying days of World War I and the black and white servicemen on their way home. In these accounts, the specific descriptions of the care for the dead, the donkey and horse carts, railway carriages, and trucks filled with linen and blanket-wrapped pyramids of corpses, often move to a shared horrified memory of witnessing, or recalling being told about, the unconscious few rising from what had been their presumed death to terrified onlookers. In every story there is a reference to the terror; the sense of common destiny; the sharing of provisions and acts of enormous selflessness; the hastily arranged food and health stations; and the desolation of survivors who lost most of their family members. Tellingly, given the lack of effective biomedical treatments at the time, the detailed descriptions of medications used and shared – such as boiled and dried garlic; boiled and mashed lemons; boiled bush remedies;

the use of iodine; carbolic acid; copious amounts of brandy, and even champagne in one account; the rubbing of tissue-salts, and the use of fresh and boiled milk, especially for pregnant women and children – also ring in unison. The accounts in this book are moving and suggestive of experience and memory knitted together as the basis for some forms of shared community for the witnesses and survivors – reminding us that the depth of this common bond could have provided the basis for a very different way out of the crisis.

Two key themes stand out. First, the movement of people (miners, traders, soldiers, workers, herbalists, wedding parties, families, school children, bankers, prisoners, farmers) speeding up through trains and new motorised transport forms, and on the water via engine-propelled ships, driving epidemics across vast territories. And second, the power of metaphor to cement memories. Phillips includes a section of black-and-white photographs of people affected by the epidemic, and of newspaper advertisements and covers of sheet music and fictional writing inspired by the 1918 'flu. In the early 1920s, Reuben Thlakele Calusa, a prodigious South African born composer, crafted several influenza-linked songs that were published by Lovedale Institution Press. The ethnomusicologist, Austin C. Okigbo, traced the lyrics and sheet music through archives and through pressed recordings of these songs made in London in the 1960s, analysing the lyrics of *Influenza 1918* that was devoted entirely to the ramifications of the epidemic in the Cape, along with the song *Intandane* (orphan).

In the first stanza of *Influenza 1918*, Caluza wrote (here translated into English by Okigbo and his researcher)[3]:

In the year nineteen eighteen

We're killed by the disease called influenza

Which finished our beloved relatives

Mothers, fathers, sisters and brothers

In other households no one was left

It took young women and men

It chose the beautiful ones

It even took the good-looking men

It took the teenagers

It took even the young maidens

It took the engaged ladies

It took the strummers [bridesmaids]

Even the grooms

It was like there was a black cloud

over the earth.

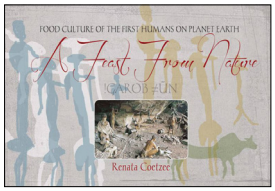
Black and grey clouds filled the mindscapes of the survivors. Angela Gilham (nee le Roux), born in Cape Town 10 years after Martha Bokako, recounted her own harrowing memories of the 'flu, in an echo of Caluza's song. She wrote a lengthy and gripping account which ended:

> *...I sincerely trust that I have not bored you with this lengthy tale but I am so thrilled to think how my own life was lengthened at that sad period whereas so many young people were sent to an untimely death. Oh! It was dreadful. Even when October comes each year and skies become grey, my thoughts still go back to 1918.*
>
> Angela Gilham (born 1901), interviewed 8 June 1972 by Richard Collier (p.29)

References

1. Phillips H, Killingray D, editors. The Spanish influenza pandemic of 1918–19: New perspective. London: Routledge; 2003.
2. Phillips H. Black October: The impact of the Spanish influenza epidemic of 1918 on South Africa. Archive Yearbook of South African History 53(1). Pretoria: The Government Printer; 1990.
3. Okigbo AC. South African music in the history of epidemics. J Folklore Res. 2017;54(1–2):87–118. https://doi.org/10.2979/jfolkrese.54.2.04

BOOK TITLE:
A feast from nature – food culture of the first humans on planet earth



AUTHOR:
Renata Coetzee

ISBN:
9780620790734 (hardcover)

PUBLISHER:
African SunMedia, Cape Town; ZAR650

PUBLISHED:
2018

REVIEWER:
Hettie C. Schönfeldt[1,2,3]

AFFILIATIONS:
[1]ARUA Centre of Excellence in Food Security, University of Pretoria, Pretoria, South Africa
[2]DST/NRF/NDP South African Research Chair in Nutrition and Food Security, University of Pretoria, Pretoria, South Africa
[3]Department of Animal and Wildlife Sciences, University of Pretoria, Pretoria, South Africa

EMAIL:
Hettie.schonfeldt@up.ac.za

HOW TO CITE:
Schönfeldt HC. Food cultures of southern Africa. S Afr J Sci. 2020;116(3/4), Art. #7555, 1 page. https://doi.org/10.17159/sajs.2020/7555

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

PUBLISHED:
26 March 2020

Food cultures of southern Africa

'Let food be thy medicine, and let medicine be thy food.' Although it is debatable whether Hippocrates (460 – c. 370 BC) actually spoke these words, scholars today agree that quality of life depends in part on the food choices that are made every day. Currently, diet is the number one risk factor in the global burden of disease[1], with malnutrition being responsible more than any other cause for poor health. Children under 5 years of age face multiple health burdens, with 22.2% stunted (low height for age) in growth, 7.5% wasted (low weight for height) and 5.6% overweight. Stunting and wasting is a strong predictor of mortality among children under 5 years and is usually the result of acute significant food shortage and/or disease. Across the globe, overweight and obesity among adults are at record levels with 38.9% of adults being overweight or obese.[2] This double burden of malnutrition and obesity ranges from the individual level to the wider society. It is possible for a person of any age to be obese and yet be deficient in important vitamins and minerals. Although agriculture provides the food that humanity requires, agricultural landscapes are becoming increasingly simplified because the variety of crops that are grown on farms is declining and is threatening agricultural biodiversity, while at the same time there is a trend towards the homogenisation of diets.

In considering current health objectives – increased life expectancy whilst being free from illness and injury with the capacity to enjoy life – as well as agriculture production worldwide, there is new appreciation for promoting and maintaining local crop varieties, indigenous animal breeds and under-utilised crop species. Harnessing our food heritage to produce more sustainable and biodiverse food for the future is steadily gaining momentum.

For this reason, the book by the late Renata Coetzee on the food culture of early *Homo sapiens* has come at an opportune time. The book is organised in five parts. The information draws on original food composition data contributed by Pinkie Wehmeyer on the nutritional value of an edible bulb (bobbejaanuintjie, *Babiana hypogea*), a flower bud (veldkool, *Trachyandra ciliata*), a seed (maramaboon, *Tylosema esculentum*) and a berry (rosyntjiebos, *Grewia flava*). Using these species, it is clear that these, and many other, veld plants have a higher nutritive content than comparable modern agricultural crops (Part 5). Currently, only 30 crops supply 95% of the energy that people obtain from food, with only four, namely maize (*Zea mays* L.), wheat (*Triticum aestivum* L.), rice (*Oryza sativa* L.) and potatoes (*Solanum tuberosum*) – supplying over 60% of calorific energy in the diet. The increase in homogenous diets and the reduction in biodiversity means that commercial agriculture may become more vulnerable to drought, pests and diseases, thus threatening food and nutrition security.[3] However, 'local' or indigenous crops, such as Cape cluster fig (*Ficus sur*), and animal breeds, such as fat-tailed sheep (Part 4), are well adapted to local conditions and are thus less reliant on chemical fertilisers, pesticides and antibiotics. These indigenous species are being replaced by a small number of so-called 'improved varieties' which are not adapted to area specific conditions and are therefore reliant on stable abiotic factors along with an array of chemicals and pesticides to deliver high yields.[4]

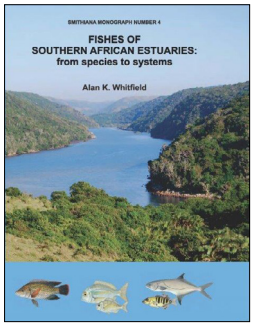
The culinary and cultural history of South African food is a subject that was close to Coetzee's heart. For many years she documented the lifestyles of past societies, the role that food played in their lives and the culture of local food consumption. In this, her final book, Coetzee explores the food culture and lifestyles of early *Homo sapiens* in Africa, hunter-gatherers and early cultivators (Part 1). Edible plant veld foods consisting of leaves, flowers, stems, bulbs, berries and fruits; as well as roots, tubers and bulbs; and bread, cakes and porridges made from some of these as harvested, preserved and prepared by Khoi-Khoin communities are recorded by Coetzee (Part 2). There is also discussion on pre-colonial protein sources that included insects, reptiles, rodents, small antelope, birds, as well as fish (Part 3).

Coetzee has combined many decades of knowledge as a nutritionist and food culture expert in this book and has brought together aspects of archaeology, palaeontology, botany, genetics, history, language and culture in a unique manner. With beautiful referenced photographs, illustrations and text, she has portrayed her content in a way that will appeal to a wider audience.

References

1. GBD 2013 Risk Factors Collaborators. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks in 188 countries, 1990–2013: A systematic analysis for the Global Burden of Disease Study 2013. Lancet. 2015;386:2287–2323. https://doi.org/10.1016/S0140-6736(15)00128-2
2. Global Nutrition Report. Global nutrition report 2018 [webpage on the Internet]. c2019 [cited 2019 Oct 27]. Available from: https://globalnutritionreport.org/reports/global-nutrition-report-2018/
3. Raneri JE, Kennedy G. Agricultural biodiversity for healthy diets and healthy food systems. London: Routledge; 2017.
4. Cook S. The spice of life: The fundamental role of diversity on the farm and on the plate. London and The Hague: IIED and Hivos; 2018.

BOOK TITLE:
Fishes of southern African estuaries: From species to systems



AUTHOR:
Alan K. Whitfield

ISBN:
9780869106397 (softcover, 495 pp)

PUBLISHER:
South African Institute for Aquatic Biodiversity, Makhanda; ZAR900 (open access ebook)

PUBLISHED:
2019

REVIEWER:
Peter B. Moyle

AFFILIATION:
Center for Watershed Sciences and Department of Wildlife Fish and Conservation Biology, University of California, Davis, California, USA

EMAIL:
pbmoyle@ucdavis.edu

HOW TO CITE:
Moyle PB. Estuaries and fishes in southern Africa: A legacy of knowledge. S Afr J Sci. 2020;116(3/4), Art. #7654, 1 page. https://doi.org/10.17159/sajs.2020/7654

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

PUBLISHED:
26 March 2020

Estuaries and fishes in southern Africa: A legacy of knowledge

As species and habitats around the world decline, the information about them expands, and is squirrelled away in an infinity of websites and obscure reports, even when collected in the interest of conservation. Yet such information, while useful, is not knowledge, which requires synthesis and understanding. Only from such knowledge, can we find better ways to understand and consequently sustain the diversity of habitats and species that make up the natural world on a broad scale. Disseminating knowledge is why books like Whitfield's are both rare and important. Whitfield has devoted much of his life to studying the biology of fishes of south African estuaries. Over 50 years he has created a rich legacy of scientific papers on the fishes, enough to satisfy most scientists. Instead, he has written a book which pulls together this diverse information in a readable fashion, the glue being his personal observations from long experience of being out on the estuaries of southern Africa sampling fish. This book allows others to tap into Whitfield's vast knowledge of estuaries and their fishes and I regard it as a model for other senior scientists to follow.

So why would anyone choose a career studying estuaries? Whitfield describes estuaries as '…regions where marine and fresh waters meet, where environmental gradients are steep, and where exceptionally high levels of production are often recorded' (p.41). Estuaries occupy very limited areas at the end of coastal rivers, where conditions vary greatly with tides, season, and freshwater inflow, making them difficult to study. Yet they are also very distinctive habitats, often important as nurseries for marine fishes and for fisheries, and their dynamic nature makes their natural history fascinating. The fishes are a mixture of marine and freshwater species that can tolerate variable salinity and temperature, with very few adapted specifically to live in estuaries. Moreover, the assemblage of fishes present at any given place is likely to change from year to year. But Whitfield makes it clear that these challenging aspects of estuaries are also what makes them so fascinating: there is always something new to learn. Not surprisingly, his book also reveals a scientist who simply loves being out sampling fish in estuaries.

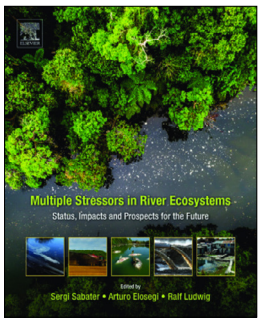
But all is not well with the estuaries of southern Africa – reflecting a global problem. They are highly vulnerable to degradation by human actions and highly vulnerable to the effects of climate change (e.g. sea level rise), which adds urgency to their study. While there are an estimated 280 estuaries in southern Africa, Whitfield calculates that 133 of them are degraded, including some of the larger ones. While there are many and multiple causes of degradation, the removal of water from inflowing rivers and alteration of catchments for farming and urban use generally lead the list. Given the short memories of people for how conditions have changed (the 'shifting baselines' phenomenon), this book makes a determined effort to compare the estuaries of today with what they were like in the past.

To increase the reader's appreciation of the fishes and estuaries, Whitfield has generated illustrations and distribution maps of 99 fish species, while 273 figures are a mixture of colour photographs and diagrams, along with 31 summary tables. The diagrams are especially attractive from my perspective (as one who also studies estuarine fishes), because many of them are accessible 'conceptual models' that present Whitfield's understanding of various aspects of the natural history of the fishes in estuaries. Figure 242, for example, shows, in colour, the existing fish community of an estuary next to a similar diagram that shows what the more diverse fish community of a healthy estuary would be like. *Anyone l*ooking at the figure can grasp what it shows, which belies Whitfield's immense knowledge behind it. The text associated with the figure provides details and references that a biologist or manager might need, especially for rating how degraded an estuary has become and what remediation measures are required. All these illustrations make the book a dream for people, like me, who like to browse through books at first acquaintance. Virtually every page has something eye-catching on it, including colour photos of the diverse estuaries of the region, which provide an outsider (also like me) an opportunity to grasp how similar (or different) the estuaries are to those in other parts of the world.

My first browse of the book revealed commonalities with the many small coastal estuaries of California, which also has a Mediterranean climate with strong seasonality of stream flows, a rich marine fish fauna, and a limited selection of freshwater fishes. This results in many similarities in how estuaries are used by diverse fishes with similar life histories. Endemic fishes can be in trouble in both places: California's delta smelt and South Africa's estuarine pipefish are both threatened with extinction because of great alterations of their estuaries. But Californian estuaries lack sea-run eels while South African estuaries lack salmon and other anadromous fishes.

The 'gee whiz' factor is high for browsers. I was struck by the photo of seven very similar species of mullet caught in the same net, which all feed on organic matter (detritus). The species apparently reduce competition by feeding in different places in the estuary because the average size of sand grains in their stomachs differs. Such observations are integrated into a larger picture to explain how the fishes use diverse estuaries, noting the roles of biological, physical and chemical processes in determining fish distribution and abundance. The result is an extraordinarily thorough exploration of estuarine ecology over a large region, which leaves as many questions as it answers – a mark of good science. Whitfield's book is a benchmark work, against which to measure the effects of global change on estuaries and fishes in southern Africa, as well as to compare these estuaries and their fishes with other estuaries around the world. Estuarine ecologists and natural historians worldwide are therefore fortunate that the book is available online at no cost.

BOOK TITLE:
Multiple stressors in river ecosystems: Status, impacts and prospects for the future



EDITORS:
Sergi Sabater, Arturo Elosegi, Ralf Ludwig

ISBN:
9780128117132 (softcover, 404 pp)

PUBLISHER:
Elsevier, Amsterdam; USD93.75

PUBLISHED:
2018

REVIEWER:
Sydney Moyo[1,2]

AFFILIATIONS:
[1]Department of Zoology and Entomology, Rhodes University, Makhanda, South Africa
[2]Department of Oceanography and Coastal Sciences, Louisiana State University, Baton Rouge, Louisiana, USA

EMAIL:
sydmoyo@gmail.com

HOW TO CITE:
Moyo S. Understanding river ecosystems in the 'Anthropocene'. S Afr J Sci. 2020;116(3/4), Art. #7772, 1 page. https://doi.org/10.17159/sajs.2020/7772

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

PUBLISHED:
26 March 2020

Understanding river ecosystems in the 'Anthropocene'

Considering the plethora of existing literature on environmental pressures facing running waters the world over, this well-written and well-illustrated book has no equal as a guide to navigating the challenges facing running waters in Africa and elsewhere, and marks an exciting step forward in discussing the management of freshwater resources in a global context. It is certain to make its mark in the canons of literature on this topic. I have already received queries from colleagues in freshwater management in South Africa and Zimbabwe as a direct result of their reading this book, yet it has only been available for a little over a year!

This first edition (hopefully not the last) of *Multiple Stressors in River Ecosystems* covers the key ecological factors and processes in rivers by dealing with the entirety of global stressors that affect running waters, whether they are small streams or large lowland rivers. The book was edited by renowned aquatic ecologist Sergi Sabater (Professor of Ecology, ICRA Catalan Institute for Water Research, Girona, Spain), stream ecologist Arturo Elosegi (Professor of Ecology, University of the Basque Country, Bilbao, Spain) and Ralf Ludwig (Professor of Physical Geography and Environmental Modeling, Ludwig Maximilian University, Munich, Germany). In addition to these esteemed editors, other extremely well-respected authors (e.g. Mažeika Sullivan, Jens Kiesel, Kelly Fouchy) were invited to broaden the thematic scope of the chapters.

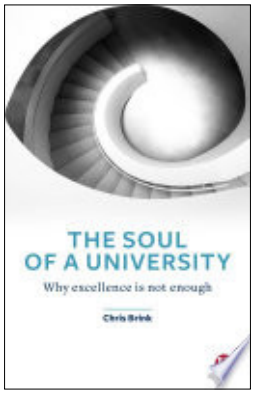
At the beginning of the book, the editors carefully explain some of the definitions that are used throughout. These definitions are useful because some terms can be easily misconstrued in freshwater ecology, for example, 'stressor', 'receptor' and 'subsidy'. The overview in Chapter 1 provides the reader with a roadmap for all subsequent chapters. Chapter 2 unravels how climate change has further compounded and modified the effect of stressors commonly occurring in running waters (e.g. land use changes). One aspect that this chapter demonstrates is just how complex it is to assess the joint effects of stressors and climate change. Viewing this contribution in the context of climate change, the reader can begin to appreciate how climate change has affected running waters in South Africa and the world over. Subsequently, Chapters 3 to 5 explore how the abiotic factors of running water ecosystems – such as hydrology and geomorphology – impact native and invasive species and how all these factors are influenced by human activities. Chapters 7 to 13 provide a thorough review of the intensity and degree of prevailing combinations of multiple stressors in different parts of the world. Chapters 14 to 20 explain models, frameworks and experimental approaches to studying multiple stressors in running waters, including the availability of drinking water to humans. Chapter 20 ends with an elegantly written integrated perspective. In the concluding chapter, Chapter 21, there is an outline of research gaps, implications for science and society, and recommendations for researchers, managers and policymakers.

Inside every chapter the arrangement of the content is excellent, but navigating through chapters can be difficult, which can be attributed to the large number of issues that are covered. I found that searching through the book could be cumbersome and time consuming; unfortunately, a glossary of important terms is missing.

My greatest praise for this book lies in the use of appropriate and interesting case studies that identify multiple stressors in riverine systems in different parts of the world (e.g. China, New World, Australia and New Zealand). Despite the strengths and interest in this book, I was disappointed to find it underwhelming in its representation of African streams. While South African researchers such as Dr Gordon O'Brien, currently at the University of Mpumalanga, contributed an overview of the stressors that affect African lotic systems in Chapter 10, I feel the book would have benefitted from additional scientists from other parts of Africa making specific contributions relating to our continent. Additionally, I would have loved the book to delve into some of the stressors that come as a result of social and political issues in management of water resources, especially considering that this work only focused on the importance of socio-economic and policy factors from a European perspective (Chapter 19). However, it is worth noting that the section on economic valuation of water resources in Europe (Section 19.5) could be extrapolated to the African continent as well.

Given the amount of work and detail that it covers, I recommend *Multiple Stressors in River Ecosystems* to graduate students and environmental managers with knowledge of the basics of riverine ecology who are interested in studying or carrying out research on the pressures affecting lotic systems in the world. To my knowledge, there is no other available book that has attempted to take a global view of stressors on riverine systems. This is a very thought-provoking read that contains the latest key scientific knowledge of the pressures facing the ecology and management of running waters.

BOOK TITLE:
The soul of a university:
Why excellence is not enough



AUTHOR:
Chris Brink

ISBN:
9781529200348 (softcover, 408 pp)

PUBLISHER:
Bristol University Press, Bristol;
ZAR350

PUBLISHED:
2018

REVIEWER:
Harry C. Boyte

AFFILIATION:
Public Work Academy, Augsburg University, Minneapolis, Minnesota, USA

EMAIL:
boyte@augsburg.edu

HOW TO CITE:
Boyte HC. The civic university and democracy's future. S Afr J Sci. 2020;116(3/4), Art. #7822, 2 pages. https://doi.org/10.17159/sajs.2020/7822

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

PUBLISHED:
26 March 2020

The civic university and democracy's future

Chris Brink, one of South Africa's leading mathematicians and a major figure in South Africa's educational transformation from apartheid, has written this important book, *The Soul of a University,* which deserves wide discussion among educators, administrators, policymakers, students and the broader public. His argument is that universities need what he calls an 'orthogonal axis' – a focus on what they are 'good for' in addressing the problems of society, in addition to the conventional assessments of what they are 'good at', i.e. disciplinary research. This type of university, which he calls the 'civic university', supports 'challenge-led research…responsive to the challenges faced by civil society, globally, nationally or regionally'. Such responsiveness requires 'civic engagement as another core function of the university', in addition to 'what they are good at' (p. 286).

To develop his case, he describes the rise and then substantial questioning, if not fall, of the 'standard model' of the research university based on the principles of academic freedom and institutional autonomy and emphasising the individual creativity of disciplinary scholars. In the standard model, academics' 'task is to question, their right is to speak, their obligation is to be objective' (p. 42). Brink has a good deal of respect for the task. He also argues that it is radically inadequate to the challenges and potentials of higher education today in societies like South Africa, which face multiplying problems.

By the 1990s, academic leaders on both sides of the Atlantic were discussing the insufficiency of the standard model. In the USA, Ernest Boyer, in *Scholarship Reconsidered,* a prestigious 1990 study for the Carnegie Foundation, proposed that the dominant view of scholarship – disciplinary research assessed by peer-reviewed publications – did not adequately describe the many functions academics need to perform in the modern university, from teaching to interdisciplinary research to engagement with society's problems. In 1994, Michael Gibbons (a theoretical physicist and Secretary General of the Association of Commonwealth Universities from 1996 to 2004) and five co-authors issued a report in the same vein, *The New Production of Knowledge* (SAGE Publishing), which addressed the need for a larger view of knowledge production than that of the standard university. 'By contrast with traditional knowledge, which we will call Mode 1, generated within a disciplinary, primarily cognitive context', they called for recognition of a 'Mode 2'. They defined Mode 2 as 'knowledge…created in broader, transdisciplinary social and economic contexts'. Mode 2, they proposed, differs from Mode 1 in a number of respects. It is transdisciplinary rather than disciplinary, heterarchical and transient rather than hierarchical, includes a diverse set of practitioners 'collaborating on a problem defined in a specific and localized context', and is best assessed by social impact (p. 46).

Brink argues that the growing challenge to the standard model was reversed by 'the rise of the rankers', international rankings like the listing of the top 500 universities in the world by Shanghai University in China, first published in 2003, followed by the *Times Higher Education Supplement* listing of the top 200 universities in the world in 2004. In today's highly competitive societies, where everything from football to hairspray is rated on scales from better to worse, rankings became the topic of intense preoccupation. As Brink puts it, 'Universities started taking note because prospective students (and their parents) had started taking note' (p. 55). 'Who's up, who's down, why are they up or down, and what does that say about the state of higher education?' moved to centre stage.

A good deal of *The Soul of a University* challenges the supposedly impartial nature of rankings based on what he calls their 'relentless linearity', as well as their negative impact on society. Linearity, Brink argues, is a way of assessing outcomes which collapses multiple dimensions into a singular continuum. It uses arbitrary measures of excellence that privilege the few and create cultures of managerial control. 'Rankings and league tables, quality as a positional good, meritocracy as rank order of worth, society stratified into classes, hierarchy instead of diversity… are all linear representations [which] compress reality into a rank list in which higher up means better and lower down means worse' (p. 227). Linearity as the singular way to measure excellence, he argues, betrays the soul of the university. His critique of university ranking systems is the finest I have seen.

Brink argues that the standard model sustains a meritocracy based on narrow understandings of 'merit', which serve neither the educational flourishing of diverse students nor the egalitarian goals of a democratic society. Today's incentives encourage the rich and well-connected to game the system, passing on their privileges to their children in ways that exacerbate economic and social inequality.

These are important arguments and Brink has the authority to make them. Over the course of his career, Professor Brink has been a fellow of the Royal Society of South Africa, President of the South African Mathematical Society, a member of the Academy of Science of South Africa, and chair of the Advisory Board of the African Institute of Mathematical Sciences. In England, where he served as Vice-Chancellor of the University of Newcastle from 2007 to 2016, he was recognised as an outstanding higher education leader. He served on the Board of the Quality Assurance Agency and the Advisory Committee on Leadership, Governance, and Management of the Higher Education Funding Council.

Especially important for South African audiences, and those abroad concerned with economic and racial equality, Brink has been a leader in making change for more inclusive higher education during the transformations from apartheid. He became Head of the Department of Mathematics and Applied Mathematics of the University of Cape Town in 1995 when the University was experiencing an influx of black students who often had no formal math background. The university was faced with the challenge of equipping them to pass the mathematics modules, mandatory for graduation in science, engineering and commerce curricula. He learned, 'what matters, in terms of both quantity and quality, is not entry but exit…the standard [students] have attained when they leave' (p 157). The mathematics department developed a variety of new pedagogies, from peer learning and the lifting of time limits on exams to a course in basic numeracy.

In 2002, Brink became Vice-Chancellor of Stellenbosch University. He writes, 'How, I wondered, does a university which had been in the forefront of apartheid become an integral part of the new multiracial South Africa?' Vision 2012, developed under his leadership, calls for the university to be 'an active role-player in the development of South African society' with 'a campus culture that welcomes a diversity of people and ideas' (p. 163). He developed impressive ways of measuring excellence that took into account context and students' narrative, not simply grades. Thus, the Rector's 'Rise Up Award', a large cash award, was based on student performance understood in context. 'When opportunity is not equally available to all, then merit cannot just be a number. The narrative, the profile, should also come into consideration.' (p. 167) The award was given at the opening of the academic year, when the vice-chancellor delivers an address to the student communities, including all new students, in the presence of senior academic staff. As each student receiving the award stepped forward, a speaker would give the story of their life and circumstances and how they were able to overcome sometimes daunting obstacles.

It is important to observe that such definitions of merit in higher education are deeply countercultural in modern society. They pose in sustained ways the question of the *ends* of education, not simply the *means* to getting to un-reflected ends. The rankings – like the current fixation on Artificial Intelligence and the Fourth Industrial Revolution – are the outgrowth of long-standing trends which the Black Consciousness scholar Xolela Mangcu has called 'technocratic creep', the spreading control by outside experts. Technocratic creep was anticipated a century ago by Max Weber, who wrote, pessimistically, about what he saw as the inevitable spread of bureaucracy and rationalisation through modern societies, 'the iron cage' (in recent translations, the 'steel carapace') of technical rationality that holds ends as constant and focuses on efficiency of means. Even more evocatively in his lecture 'The Profession and Vocation of Politics', Weber described this dynamic of instrumental rationality as 'the polar night of icy darkness'.[1(p.368)] Technocratic cultures, driven by efficiency dynamics, take ends to be a given, whether rankings, winning elections, test scores in lower grades, profits in businesses, or service delivery to citizens conceived as customers. They are accelerated by the digital revolution.

These are international challenges. In 1997, the Kellogg Foundation, an American philanthropy, asked the Center for Democracy and Citizenship which I directed at the University of Minnesota, to assess whether what is called the 'land grant mission' – the civic purpose of what are known as land grant colleges and universities – could be renewed. Edwin Fogelman, Chair of Political Science at the University, and I interviewed several dozen senior faculty members with distinguished reputations in different fields about their work experiences. We found hidden discontent with the increasingly competitive research culture of the university. Almost all disliked the 'star' system, in which well-known professors seek outside offers to enhance their salaries. They were dismayed about the erosion of cooperative and interdisciplinary work and the devaluation of undergraduate teaching. Most to the point, many expressed anguish about silencing of the discussion of public purpose by the norms of detachment which had come to permeate the university. The Center worked with the Provost of the university, roughly equivalent to a vice chancellor, to create a cross-university Civic Engagement Task Force charged with developing strategies for strengthening the public purpose of scholarship, teaching and other forms of professional work. It generated many innovations. We also came up against the iron grip of the national and international rankings. At one of our symposia we debated the conflict between rankings and civic purpose. Professor Gail Dubrow, Dean of the Graduate School, remarked that the university, if it aspired to greatness, faced a choice between two alternative paths. It could seek to 'play the game' of competing with universities in global ratings (the University of Minnesota is currently 79th in global rankings[2]). Or it could work with other universities and partners in the larger society to change the game.[3,4]

South Africa is a world-class example of the struggle for human freedom with values of the anti-apartheid struggle enshrined in a constitution admired by democracy advocates across the world. It warrants a system of higher education that lives up to this example. It will also take an international movement to change the game. Chris Brink has written a book that contributes to the process.

References

1. Lassman P, Speiers R. Weber: Political writings. Cambridge, UK: Cambridge University Press; 1994.
2. The World University Rankings. University of Minnesota [webpage on the Internet]. c2020 [cited 2020 Jan 21]. Available from: https://www.timeshighereducation.com/world-university-rankings/university-minnesota
3. Boyte HC. Everyday politics: Reconnecting citizens and public life. Philadelphia, PA: University of Pennsylvania Press; 2004.
4. Boyte H. Public engagement in a civic mission. Dayton, OH: Kettering Foundation Press; 2000. Available from: https://www.kettering.org/catalog/product/public-engagement-civic-mission-case-study

AUTHOR:
Alexander Quandt[1]

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

CORRESPONDENCE TO:
Alex Quandt

EMAIL:
alex.quandt@wits.ac.za

HOW TO CITE:
Quandt A. All materials great and small. S Afr J Sci. 2020;116(3/4), Art. #7906, 4 pages. https://doi.org/10.17159/sajs.2020/7906

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
materials science, multi-scale modelling, 2D materials, quasicrystals, topological materials, plasmonics

PUBLISHED:
26 March 2020

All materials great and small

Professor Alexander Quandt is the recipient of the 2018/2019 NSTF-South32 Special Annual Theme Award: Materials for Inclusive Economic Development (in recognition of the United Nations 'International Year of the Periodic Table of Chemical Elements').

One of the oldest arts of humankind is toolmaking. Over the millennia, all civilisations have experimented with different types of raw materials. Many generations of stonemasons, blacksmiths and gaffers have invented better and better types of nuts and bolts to fix old things, and to facilitate the making of new things. These are the unknown pioneers upon whose shoulders our modern civilisation stands today, and their forgotten contributions to science and engineering also mark the origins of materials science.

If toolmaking is so deeply embedded in the DNA of our species, then our desire to understand the secrets of the universe around us appears to be another major driving force. Through this desire, we have essentially learned two important lessons. First, that we need extremely sophisticated technical instruments to explore the world around us at the smallest and the largest length scales. And second, that the knowledge gained from the study of the physical processes at these extreme scales pales in comparison to the complexity and the sophistication of even the simplest version of the molecular nanomachineries which keep us 'alive and kicking' on a day-to-day basis.

Colleagues who are not from the field of materials science usually tend to dismiss such remarks by pointing out that the complexity of living organisms stems from the peculiar nature of a handful of light elements which are at the heart of organic chemistry. Inorganic materials by contrast are generally perceived to be as boring and as well understood as rock salt, and thus so must be the whole field of materials science. Nothing could be further from the truth.

I hope that through this Commentary, and by talking about some of the highlights I have encountered during a long career in computational materials science, I will be able to convince readers that modern materials science amounts to very much more than serving as the warehouse clerks in the supply chain of industry and other scientific fields.

Nature of the chemical bond

The year 2019 marked the 150th anniversary of the periodic table of chemical elements in its most popular form, first created by Dmitri Mendeleev. The filling of the missing gaps in the periodic table was anything but a civilized and courteous endeavour – the bitter race to fame and worldwide media attention that we often witness today is obviously something that was already common in the 'good old days'.[1] But despite all the battles, the discovery of the chemical elements was just the beginning of a long journey into the modern science of materials – a journey which is far from over.

One highlight provided by the periodic table was the prediction of a striking chemical similarity between some of its elements; a similarity which is based on the shell structure of the electrons in an atom – a fact that could be explained to some extent using simple quantum mechanics.[2] From that point onwards, scientists knew at least something about the quantum mechanical foundations of materials science. The next major step was to find out how these various atoms would combine, which amounted to a systematic description of the atomic structure of organic and inorganic materials.

This task was taken up by eminent scientists like Linus Pauling. In his famous book[3] *The Nature of the Chemical Bond* (first published in 1939), Pauling essentially presented the first systematic description of chemical bonding in materials. His main resources were crystallographic data and a very intuitive approach to quantum mechanical many-electron systems. His elegant style of combining experimental data with basic quantum mechanical concepts had a deep impact on generations of computational materials scientists, including myself. When I first stumbled over a well-worn copy of Pauling's book as a high school student, I was astounded by the complexity of some of the inorganic materials described in his book. That was the moment when I decided to study physics and chemistry.

It was a long way from Pauling's initial ideas to the almost industrial-scale type of computational infrastructure that aids today's search for new materials. A key method that drove the development of computational materials science was density functional theory (DFT), which was developed by Walter Kohn and co-workers in the late 1960s.[4] Based on this novel and groundbreaking many-particle method, it was possible to predict the key materials properties of materials with high accuracy and with reasonable numerical resources. The numerical implementation of DFT required many subtle tricks of the trade, and the article by Payne et al.[5] in *Reviews of Modern Physics* (1992) was a milestone in this development. Almost all the popular programme packages in use today are based on at least some of the numerical methods described in this detailed review article.

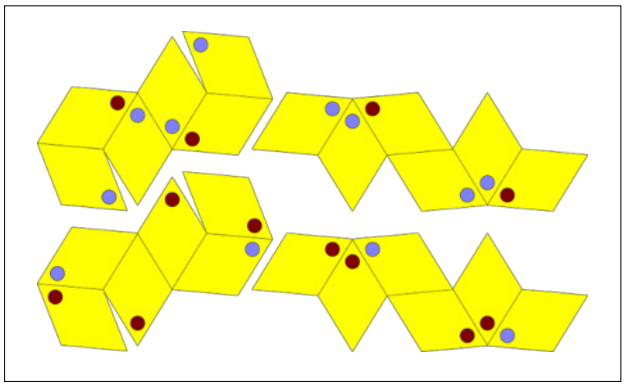
One of the co-authors of this influential article is Mike Teter, who was my postdoctoral supervisor at Cornell University (USA). He taught me that computational materials science is a great deal more than pushing buttons on computational black boxes. If done properly, it also involves a constant improvement and optimisation of the underlying numerical methods. A survey of the latest many-particle simulation methods and their implementation on supercomputing facilities can be found in another monograph[6], *Computational Many-Particle Physics*, to which I contributed two chapters. This book has been downloaded from Springer so many times that a copy of it must sit on nearly every computer running materials simulations today.

Surprises

If you are late for a conference dinner and are seated at a table with colleagues that you hardly know, you might want to break the ice by starting a typical scientist's joke: 'A crystallographer, a physicist, a chemist, a mathematician,

a materials scientist and a mineralogist find themselves in a room…'. However, you would not have much success with such a joke at a typical quasicrystals conference, because such is typically the composition of these conferences. Since the discovery of quasicrystals by Dan Shechtman and co-workers in 1984,[7] these materials have put into question some of the most fundamental concepts in crystallography and condensed matter physics. Among them is the question of whether a material with no obvious translational order but sharp X-ray diffraction peaks would still qualify as a crystal or not. (Answer: it does!). But sadly, some of the godfathers of modern materials science like Linus Pauling dismissed the field of quasicrystals at the outset.[8]

In hindsight, the borderline existence of the field of quasicrystals among the condensed matter establishment turned out to be a blessing. In the relative tranquillity of the quasicrystals conference, a very diverse scientific community met and worked out groundbreaking new ideas about the structure and the properties of complex periodic and aperiodic materials. My PhD supervisor, Peter Kramer, was among the pioneers of this field. He developed one of the first tiling models for icosahedral quasicrystals.[9] In Figure 1 I have posted some instructions to fabricate paper versions of the corresponding tiles, and as this is a Commentary about toolmaking, you are invited to produce paper versions of the corresponding tiles, and try to tile space in an aperiodic fashion, but without creating holes or overlaps between those tiles.



Source: Wikipedia

Figure 1: Unit cells of a quasi-periodic icosahedral tiling. Transfer these templates to cardboard and enjoy tiling.

Under the supervision of Professor Kramer I started to develop some of the first atomistic models for layered and icosahedral quasicrystals, one example being the Quandt–Elser model for AlPdMn.[10] Today, the combination of experimental or crystallographic probes with DFT-based materials simulations has become mainstream in the field of quasicrystals research.

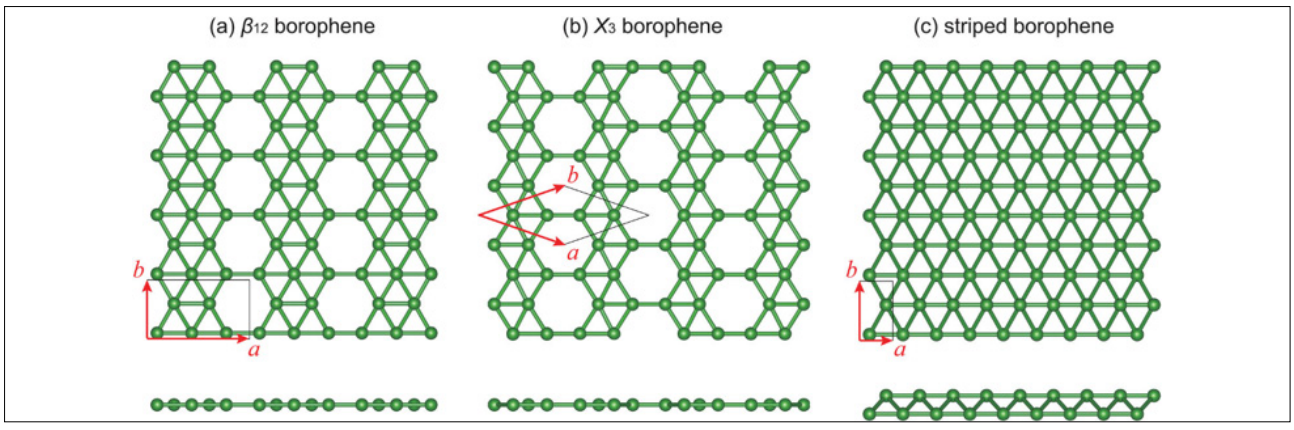
Another case in which Pauling found himself corrected was the case of boron. In his famous book[3] he goes to great lengths to explain the astonishingly complex nature of the chemical bond in the elementary phases of boron and makes a strong case for boron icosahedra as the building blocks of such materials. In a collaboration with my colleague Ihsan Boustani, we took this idea very seriously and tried to develop a structure model for a one-elemental boron quasicrystals built on Pauling's ideas. To our surprise we discovered something completely different: namely that boron forms quasi-planar clusters rather than icosahedra, which gives rise to boron nanotubes, boron fullerenes and quasi-planar boron sheets (now called borophene).[11] Several allotropes of borophene are shown in Figure 2.

The formulation of an Aufbau principle for boron clusters marked the birth of the field of boron nanomaterials.[11] Historically, the prediction of quasi-planar boron sheets predated the discovery of graphene by several years. Unfortunately, borophene was much harder to synthesise than we originally thought it would be, whereas its cousin material, graphene, has become one of the biggest sensations in the story of materials science.[12] The only materials class in which carbon cannot really compete with boron is the possible existence of boron quasicrystals.[11] Up until now, however, the boron quasicrystal remains our only prediction which has not been confirmed over the years. But maybe something as simple as a pure sample of alpha-boron and a ball mill might do the job?

Applied computational materials science

One aspect of cutting-edge computational materials science is the systematic and almost industrial scale search for new materials.[13] This seems to be a promising strategy, provided one is privileged enough to gain very generous access to the corresponding supercomputing infrastructure. Another promising approach to computational materials science is the modelling of possible technological devices. Note that atomistic methods like DFT can only be applied to idealised model systems comprising up to several hundred atoms. This number falls short by at least 20 orders of magnitude compared to the number of atoms forming a typical solar cell or nano-optical device. The way out of this dilemma is the use of multiscale approaches,[14] in which device simulations are essentially based on phenomenological models, but the key materials parameters for these phenomenological models may be taken either from experiment or from an atomistic simulation using DFT. This allows for 'in silico' development of new types of devices without ever going to the laboratory. Over the years we were able to demonstrate that this approach works extremely well in the field of photovoltaics.[15]

Another field in which we had explored a full suite of fundamental and phenomenological simulation tools is computational plasmonics.[16] Plasmonics is a key optical technology based on surface waves, which is supposed to bridge the fields of electronics and optics. Figure 3 shows predictions of plasmonic surface waves ('plasmons') for graphene and borophene made by my colleague Robert Warmbier, which is a follow-up on our earlier work referred to above. The strongest peaks in the spectra



Source: Wikipedia

Figure 2: Allotropes of borophene.

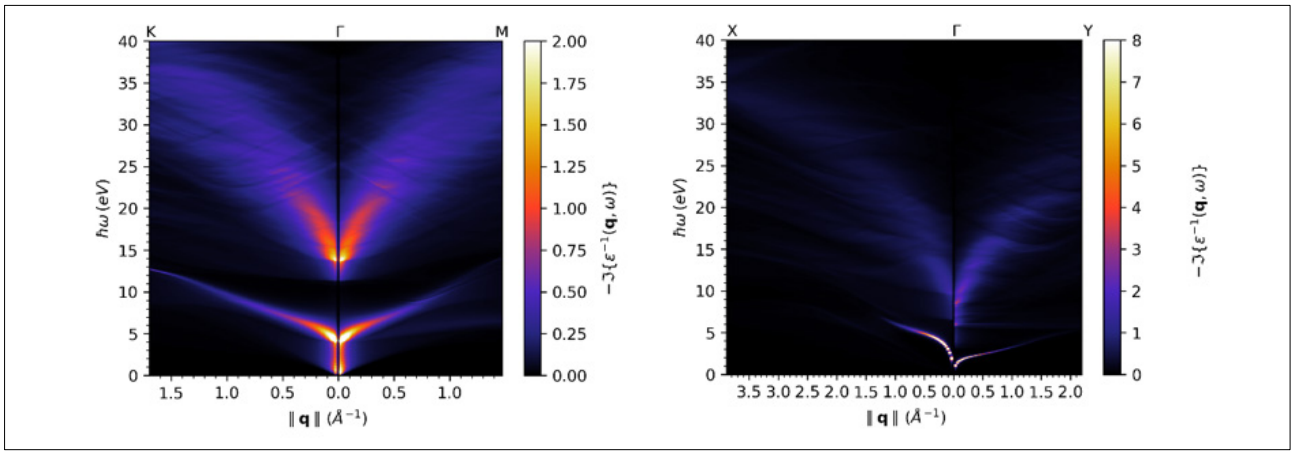


Figure 3: Simulation of Electron Energy Loss Spectrum (EELS) for graphene (left) and borophene (right). Note the strongest resonance peaks, which predict THz plasmons for graphene (1–100 meV), and visible range plasmons for borophene (1–3 eV).

correspond to resonances induced by surface plasmons. For graphene we notice the existence of plasmons in the UV (beyond 3 eV) and in the THz range (1–100 meV), whereas some of the strongest peaks seen in the case of borophene indicate the existence of plasmons within the optical range (1–3 eV).

Even more surprises

For a born toolmaker, even the most exotic natural phenomena can be easily understood after referring to a material or a process that the toolmaker could move with their own hand. I was rather surprised to find out from a remarkable book by Strogatz[17], that undisputed millennial geniuses like Archimedes and Newton did not make their biggest mathematical discoveries based on abstract reasoning but on simple mechanical analogues.

It turns out that materials science provides us with a fascinating zoo of possible analogies to particles and processes, which are otherwise found only at some of the smallest and at some of the largest length scales in our universe. This was the topic of a rather visionary book by Volovik[18]. It almost appears as though materials science provides us with something like the mysterious aleph, which has so vividly been described in the famous short story by Borges[19].

As it happened, Volovik was immediately vindicated by the discovery of topological insulators and related materials[20], which brought into materials laboratories worldwide effects and concepts that previously were only known from field theory and high energy physics. Based on these materials analogies, we now have the unique possibility of studying model versions of black holes, magnetic monopoles and new elementary particles in a materials lab, and hopefully will learn a good deal about their fundamental properties before we start to hunt for the same physics elsewhere using giant telescopes and particle accelerators.

Summary and outlook

It is the complex nature of the chemical bond that gives rise to an amazing variety of materials, some of which have been described in this Commentary. Among them we will surely find new classes of 'wonder materials' which will drive the technologies of the future. Over time, these new materials will lead to a fundamentally different portfolio of basic industrial materials. When this happens, countries like South Africa, which have a wide range of minerals resources, will be in a very comfortable strategic position. However, these mineral-rich countries will need to use their natural resources wisely, and they must also succeed in building a strong beneficiation industry around their mining sectors and include investments in the necessary human capital.

Some steps in that direction have already been made in South Africa by funding internationally established institutions like the Centre of Excellence in Strong Materials (CoE-SM) with headquarters at the University of the Witwatersrand. Beyond these national efforts, there are also new initiatives like the Centre of Excellence in Materials, Energy and Nanotechnology (CoE-MEN), which aims to provide materials scientists in Africa with a new continent-wide research platform, and which operates under the auspices of the African Research Universities Alliance.

These initiatives will need time and continuous financial and technical support to evolve to their full potential, and one should not make the mistake of expecting that the futuristic Wakanda of the Marvel Universe will arise overnight. But in the CoE-SM and the related Materials for Energy Research Group, we have already seen many cases in which teams of researchers start from the atomistic simulation of the key materials and end up with the production of sophisticated high-tech devices.

And so, the story of toolmaking will continue, and new discoveries will surely surprise us on our long and winding road to 'Wakanda'.

Acknowledgement

I thank Prof. Jane Carruthers, Editor-in-Chief, for the opportunity to write this personal essay on materials science.

References

1. Scerri E. A tale of seven elements. Oxford: Oxford University Press; 2013.
2. Slater JC. The theory of complex spectra. Phys Rev. 1929;34:1293. https://doi.org/10.1103/PhysRev.34.1293
3. Pauling L. The nature of the chemical bond. 3rd ed. Ithaca, NY: Cornell University Press; 1960.
4. Kohn W, Sham LJ. Self-consistent equations including exchange and correlation effects. Phys Rev. 1965;140:A1133. https://doi.org/10.1103/PhysRev.140.A1133
5. Payne MC, Teter MP, Allan DC, Arias TA, Joannopoulos JD. Iterative minimization techniques for ab initio total-energy calculations: Molecular dynamics and conjugate gradients. Rev Modern Phys. 1992;64:1045. https://doi.org/10.1103/RevModPhys.64.1045
6. Quandt A. Ab initio approach to the many-electron problem; ab initio methods applied to structure optimization and microscopic modelling. In: Fehske H, Schneider R, Weisse A, editors. Computational many-particle physics. Berlin: Springer; 2008. p. 415, 437. https://doi.org/10.1007/978-3-540-74686-7_15
7. Shechtman D, Blech I, Gratias D, Cahn J. Metallic phase with long-range orientational order and no translational symmetry. Phys Rev Lett. 1984;53:1951. https://doi.org/10.1103/PhysRevLett.53.1951
8. Cahn JW, Gratias D, Shechtman D. Pauling's model not universally accepted. Nature. 1986;319:6049. https://doi.org/10.1038/319102a0
9. Kramer P, Neri R. On periodic and non-periodic space fillings in E ^ m obtained by projection. Acta Crystallogr. 1984;A40:580. https://doi.org/10.1107/S0108767384001203
10. Quandt A, Elser V. Ab initio based modeling of i-AlPdMn. Phys Rev. 2000;B61:9336. https://doi.org/10.1103/PhysRevB.61.9336
11. Quandt A, Boustani I. Boron nanotubes. ChemPhysChem. 2005;6:2001. https://doi.org/10.1002/cphc.200500205

12. Geim AK. Graphene: Status and prospects. Science. 2009;324:1530. https://doi.org/10.1126/science.1158877

13. Saal JE, Kirklin S, Aykol M, Meredig B, Wolverton C. Materials science and discovery with high-throughput density functional theory: The open quantum materials database (oqMD). JOM. 2013;65:1501. https://doi.org/10.1007/s11837-013-0755-4

14. Weinan E. Principles of multiscale modelling. Cambridge: Cambridge University Press; 2011.

15. Quandt A, Mokgosi I, Warmbier R. Simulations of conventional and augmented types of solar cells. In: Enrichi F, Righini G, editors. Solar cells and light management. Amsterdam: Elsevier; 2020. p. 249. https://doi.org/10.1016/B978-0-08-102762-2.00007-0

16. Mohammed F, Warmbier R, Quandt A. Computational plasmonics: Theory and applications: Numerical techniques. In: Agrawal A, Benson T, De La Rue RM, Wurtz GA, editors. Recent trends in computations photonics. Springer series in optical sciences. Cham: Springer; 2017. p. 315; 341. https://doi.org/10.1007/978-3-319-55438-9_12

17. Strogatz S. Infinite powers. New York: Houghton Mifflin Harcourt; 2019.

18. Volovik GE. The universe in a helium droplet. Oxford: Oxford University Press; 2003.

19. Borges JL. The aleph. London: Penguin Books, 1998. Reprint from 1949.

20. Hasan MZ, Kane CL. Colloquium: Topological insulators. Rev Modern Phys. 2010;82:3045. https://doi.org/10.1103/RevModPhys.82.3045

AUTHOR:
Nithaya Chetty[1]

AFFILIATION:
[1]Faculty of Science, University of the Witwatersrand, Johannesburg, South Africa

CORRESPONDENCE TO:
Nithaya Chetty

EMAIL:
Nithaya.Chetty@wits.ac.za

HOW TO CITE:
Chetty N. Re-inventing inventiveness in science. S Afr J Sci. 2020;116(3/4), Art. #7824, 2 pages. https://doi.org/10.17159/sajs.2020/7824

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
science advancement, postgraduate training, innovation, science for society

PUBLISHED:
26 March 2020

Re-inventing inventiveness in science

Much scientific research is mundane, often involving only incremental advancements to our understandings. Sometimes progress is marginal, for example when established methods are applied to different systems or when ideas and techniques are improved slightly. In this way, advances often come at a painstakingly slow rate, and not always in a linear fashion. Scientific dead ends are a part of the journey to scientific advancement.

Some will argue that it seems a luxury for scientists to be rewarded for unfettered research when their individual contributions to the overall advancement of the field of inquiry can be described as miniscule, with little or no direct impact on society or on commerce and industry. Given the enormous expense involved in supporting research at universities and our national facilities, especially in a developing country like South Africa, it seems fair to ask: Why should society, more specifically taxpayers, support open-ended scientific research?

Free open-ended inquiry as a basis for scientific advancement

This question has been dealt with in different eras by different societies, and many industrialised nations have long seen the necessity for nurturing a community of scientific specialists. It is not necessarily the individual contributions that count, but the aggregated understandings that can lead to important scientific advances. If we are going to progress, South Africa needs to be part of this journey.

At times, however, bright new ideas, incisive innovative theories and explanations, radically different techniques and approaches, ingenious inventions, unpredicted observations, and unimagined discoveries emerge. This type of scientific breakthrough often cannot be planned in advance nor can it be predicted. Nor can it be directed from the top, or orchestrated by managers and bureaucrats, although the support of these groups is essential. Scientific breakthroughs thrive where intellectual enquiry is supported and free.

Scientific advancement as a basis for directed, applied industrial research

Unfettered science may seem to be divorced from South Africa's practical problems. Should we, therefore, move in the direction of more directed, applied industrial research – by which I mean research focused on solving a known set of problems with direct benefit for commerce, industry, or society more generally – to address societal challenges?

There is a strong argument that directed, applied industrial research focused on practical problem-solving needs strong support, especially in a country like ours with high levels of poverty and unemployment. However, this research should not be done at the expense of free, open-ended inquiry. For if it did, we would run short of truly fresh scientific ideas to apply to our societal problems. We would destroy the source of high-quality young graduates with the capability and versatility to move towards directed, applied industrial research and thus we would destroy the entire scientific enterprise.

Quality postgraduate student training as a vital cog for societal development

The South African government has stated its aim to graduate 6000 PhDs per year by 2025. Currently, we produce about half this number annually across all disciplines. Although currently unrealistic, the target might be regarded as aspirational, but it also emphasises postgraduate student development as an instrument for societal change. (Here, by postgraduate students I mean the cohort of PhD students.)

It is important that we focus on quality postgraduate student development, and that we re-double our efforts to enable graduates to lead successful careers while making meaningful contributions to society. Unless we plan properly, this will not happen. Because the vast majority will not find academic employment, as our universities and research facilities have their own fiscal constraints, the need to train postgraduates for success beyond academia is extremely urgent.

Training postgraduate students in an environment of open-ended inquiry is vital for their personal development. This is where critical thinking is nurtured, and students require intellectual versatility. They must learn skills transferable to different employment settings. Thus, the question arises about how we can be more creative in postgraduate student training so that graduates can impact more positively on society without damaging the very ethos needed to sustain science into the future.

Unless we grapple with this question, we may run the risk of losing societal support for science. Having thousands of unemployed or underemployed PhDs will come back to haunt South African academia if we do not address this question timeously. Increasingly, it will be seen to be a failure if our postgraduates are not sufficiently inventive and if they are largely incapable of creating employment opportunities, at the very least, for themselves.

The following are, in my view, some of the principal areas in which we might re-orientate postgraduate student training to avoid what appears to be a looming future crisis of massive underemployment of highly qualified postgraduates.

1. Enhance critical thinking and the ethical practice of science

With ready and free access to information on the Internet, are our universities, and hence our graduates, becoming less relevant? It is likely that this will happen if we do not adjust our research, teaching and learning programmes. Despite the deluge of information, students still need to be taught the broad understanding of their subject material and assisted in making inter- and multi-disciplinary connections. They need to engage critically with their subjects and be trained to become more discerning about 'information'.

The ethical practice of science for humanity must become a very strong focus of our training across all scientific disciplines if postgraduate students will be more relevant for society. Society needs more scientists with a more humane view of this world.

2. Strengthen scientific and technical skills development

In many experimental sciences, the focus is on using off-the-shelf commercial equipment. While this often leads to excellent scientific outcomes, there is often little attention to innovating with the instrumentation itself. Thus, many postgraduate students become operators of equipment rather than its designers, modifiers or even developers. This impacts negatively on skill sets, often with reduced options for employment outside academia.

This reality replicates itself in many different scientific endeavours. For instance, in the fields of computational sciences a generation ago, students and postdoctoral fellows would have developed their own codes incrementally over many years. Today, many production codes have been commercialised with exorbitant licence fees, although the community of open-source developers is growing. These codes are highly specialised and sophisticated and written by teams of experts. It is obviously impossible for a postgraduate student to write an entire production code that can compete with what has been professionally developed. But we need students to be computationally competent if they are going to seek work in commerce and industry.

The reality is that these large-scale production codes are being used as 'black boxes' by postgraduate students and researchers alike. Students do not always learn computational theory and algorithms in a fundamental way, and many can barely make substantial modifications to the codes. Worse still, in the cases of most licensed software, the source codes are not available for scrutiny. The consequence is that science is impeded in very real ways. The question we should ask is: Are we producing computational scientists or computational operators?

The irony is that advances in science and technology, especially in automation – while important for society-at-large – have helped dull the creativity and inventiveness of our postgraduate students, with bleak consequences for their futures, and possibly for science itself. We need to fix this.

There is new science to discover if the limits of instrumentation can be pushed to new levels. Supervisors need to take a more pedagogical approach to training postgraduate students in using experimental equipment. Postgraduate students from all branches of science need to be capable of mechanical machining and electronic design. They should learn to be able to dissect equipment and discover its inner workings, to modify it and even design and build new capabilities

All students, including those in the humanities, should be able to write computer code. How else will they be able to translate a new theoretical model that they might have developed as a part of their research into tangible results?

We must impart to our postgraduate students the types of experimental, technical and computational skills that will enable them to succeed outside academia.

3. Exploring the path from science to innovation

Researchers are often accused by managers, funders and society in general of not being sufficiently innovative. There is growing pressure for universities to engage more with discovery and innovation. However, I have already argued above that open-ended unfettered research must continue to be the bedrock of our university research systems. Are we going to continue to live in separate worlds, or can we look for ways in which we can bridge this growing divergence?

The current enterprise system at many universities works as follows. When a new idea is discovered, it is identified and then advances along the long and arduous path toward commercialisation. A new set of skills and understanding is critical for this to happen, often totally unknown to the typical scientific researcher. For example, major funding is almost always vital; prototyping is necessary; extensive market research is needed; clinical trials may be required; intellectual property rights require investigating; patenting is expensive; etc.

This chain pre-supposes that new, commercially viable ideas are actively pursued in the research laboratories. This, however, is often not the case as many academics focus on their academic pursuits and do not actively seek marketable ideas. And even if academics stumble across an exciting new and marketable idea, they do not often have the time to take it further.

However, the situation is very different for postgraduate students who are not faculty members. The requirement for research papers of international standing, and a thesis, remain important requirements, but students need to be exposed to a different way of thinking about research. Research can lead to new ideas and discovery, but innovation reflects the chain of processes to bring this research to the market or in service of society, which, as was noted above, requires other knowledge and skills to which postgraduate students are not generally exposed.

We should educate and train postgraduate students to take innovative ideas through to market, by including practical real-life examples taught by successful experts, from the outset, rather than as an option to be considered only at the end when a student fortuitously finds an exciting idea that they wish to pursue. This cannot be left to chance.

We need a new conversation: How can we more effectively expose our postgraduate students to the ideas of innovation at the start of their studies rather than accidentally at the end? Inspiration can change attitudes. Not all PhDs will become entrepreneurs, but after such exposure, they will always seek the value of their work in terms of application.

The future of the South African academy depends on training an increasing number of highly qualified postgraduates who can readily be employed outside academia, or who can create their own employment. Major public funding and effort will be devoted to increasing our postgraduate student production rate as intended. If this is not accompanied by increasing quality and employability, then we are heading for catastrophe.

By focusing on the employability of our postgraduate students, open-ended unfettered research – which I have argued is essential for the long-term viability of our scientific research systems – must be able to continue in an unhindered way. The re-orientation of postgraduate student training along the above lines will enhance the quality for open-ended research and increase the attractiveness of postgraduate studies. The top 5% of PhDs may still find excellent academic jobs, but I expect and hope that they will, through the process I have outlined, have developed a greater understanding and sympathy for the plights facing the remaining 95%.

Acknowledgement

I thank Dr Neville Comins, former Director of the Innovation Hub, for useful discussions.

AUTHORS:
Michael D. Fraser[1]
Bruce A.S. Henderson[2]
Pieter B. Carstens[2]
Alan D. Fraser[1]
Benjamin S. Henderson[2]
Marc D. Dukes[2]
Michael N. Bruton[3]

AFFILIATIONS:
[1]Citizen Scientist, Pumula, KwaZulu-Natal South Coast, South Africa
[2]Citizen Scientist, Somerset West, Western Cape, South Africa
[3]Honorary Research Associate, South African Institute for Aquatic Biodiversity, Makhanda, South Africa

CORRESPONDENCE TO:
Mike Bruton

EMAIL:
mikefishesbruton@gmail.com

HOW TO CITE:
Fraser MD, Henderson BAS, Carstens PB, Fraser AD, Henderson BS, Dukes MD, et al. Live coelacanth discovered off the KwaZulu-Natal South Coast, South Africa. S Afr J Sci. 2020;116(3/4), Art. #7806, 3 pages. https://doi.org/10.17159/sajs.2020/7806

ARTICLE INCLUDES:
☐ Peer review
☒ Supplementary material

KEYWORDS:
Latimeria chalumnae, African Coelacanth Ecosystem Programme

PUBLISHED:
26 March 2020

Live coelacanth discovered off the KwaZulu-Natal South Coast, South Africa

When a living coelacanth was trawled off East London, South Africa, at a depth between 72 m and 100 m on 22 December 1938, it caused an international sensation.[1-7] The specimen was saved for science by the young curator of the East London Museum, Marjorie Courtenay-Latimer[8] and identified by JLB Smith of Rhodes University College[5,6]. Smith named it *Latimeria chalumnae*, after Courtenay-Latimer and the river off which it was caught, and it is listed in the official Coelacanth Conservation Council (CCC) inventory as CCC 1. *L. chalumnae* is classified in the family Latimeriidae and the suborder Latimerioidea.

JLB Smith, who was a keen angler and had an excellent knowledge of fish anatomy, behaviour and habitat preferences, predicted that the East London fish was a stray from relatively shallow rocky reefs further north on the tropical east coast of Africa.[7] Rumours of dead coelacanths being washed up on the seashore at Knysna and Gonubie near East London, or sighted by divers in Mozambique, have not been substantiated.[9] In 1989, Bruton[10] predicted that coelacanths would be found off the Transkei and Zululand coasts but this suggestion was rejected at the time.

L. chalumnae has since been found in the Comoros (over 200 specimens)[11], Mozambique (one specimen, CCC 162)[12], Madagascar (over 13 specimens)[13], Kenya (one specimen, CCC 178)[14] and Tanzania (over 70 specimens)[15,16], and a colony was discovered in the isiMangaliso Wetland Park in northern Zululand, South Africa, in 2000[17]. Another species of living coelacanth, *L. menadoensis*, was found on the other side of the Indian Ocean off North Sulawesi Island in Indonesia in 1997 (CCC 174)[18,19]; this species is smaller than *L. chalumnae* and its body, which is also covered with white spots, is brown rather than blue.

The first scientist to observe a living coelacanth was Jacques Millot of France who briefly examined a dying immature female fish (142 cm, 41 kg; CCC 8) in a flooded wooden boat at Mutsamudu on Anjouan Island in the Comoros in 1954.[20] The fish was captured at 20:00 on 12 November 1954 and was kept alive in the sunken boat from about 23:30 until 15:30 on 13 November 1954. The fish was stressed and exhibited only feeble movements.

Several coelacanths that were subsequently caught by traditional fishermen in the Comoros (mainly Grand Comoro), and brought to the attention of scientists, survived for periods of 1–42 h, usually 1–11 h, near the water surface where they could be observed by divers.[9]

Hans Fricke and his team from Germany were the first to study the living coelacanth in detail from their research submersible *Jago.* They compiled an extraordinary data series on the living coelacanth in the Comoros spanning 21 years and including 145 specimens that had been individually identified using the unique patterns of white spots on their bodies.[21-26]

On 28 October 2000, mixed-gas divers Pieter Venter, Peter Timm and Etienne le Roux discovered coelacanths living at a depth of 104 m in Jesser Canyon at Sodwana Bay[17] in the newly proclaimed iSimangaliso Wetland Park in Maputaland – the shallowest sighting of coelacanths at that time. On 27 November 2000 they filmed three coelacanths at a depth of 106 m in Jesser Canyon. These discoveries led to the establishment of the African Coelacanth Ecosystem Programme (ACEP) in April 2002 which aimed to initiate and promote a new phase of multidisciplinary research on the coelacanth and its habitats.

The South African Institute for Aquatic Biodiversity in Makhanda (previously Grahamstown) was appointed as the lead organisation for ACEP, which has been carried out in three phases: 2001–2006, 2007–2011 and 2012–2015. From 2002 to 2004, Professor Hans Fricke and his team returned to South Africa with the *Jago* submersible to study coelacanths in the iSimangaliso Wetland Park from the R/V *Algoa* as part of ACEP. They carried out 47 survey dives with a total bottom time of 166 h at depths ranging from 46 m to 359 m. Initially, 24 coelacanths were identified in three submarine canyons at depths from 96 m to 133 m along a 48-km stretch of coast in the iSimangaliso Wetland Park. This number was later increased to 32 individuals.[27,28] Over time the ACEP programme extended its research programme further north into other countries in East Africa and the Western Indian Ocean Islands using the R/V *Algoa* and other motherships, the *Jago* submersible and a *Sea-Eye Falcon* underwater remotely operated vehicle.

On 15 February 2004, mixed-gas diver Christo van Jaarsveld observed a coelacanth at a depth of about 54 m in Diepgat Canyon south of Sodwana Bay in the iSimangaliso Wetland Park (2004, personal communication to Mike Fraser, February 16; confirmed 26 November 2019). This sighting is the shallowest on record for a healthy adult coelacanth and was the 19th specimen known from the iSimangaliso Wetland Park, but it has not been seen again. Van Jaarsveld has subsequently seen and filmed another specimen on 7 August 2018 in the iSimangaliso Wetland Park (2018, personal communication to Mike Fraser, August 7); this fish has also not been seen since.

At about 09:00 on 22 November 2019, a team of divers observed and filmed a single coelacanth (ACEP no. 34) at a depth of 69 m off the village of Umzumbe (between Hibberdene and Pumula) on the South Coast of KwaZulu-Natal (Figure 1). This site is about 325 km south of the iSimangaliso Wetland Park. The divers, Mike Fraser and Alan Fraser from Pumula and Bruce Henderson and Pieter Carstens from Somerset West, launched from the Injambili launch site at Pumula, with Benjamin Henderson and Marc Dukes acting as surface support in the boat. Henderson and Carstens used open circuit trimix and Mike and Alan Fraser used rebreathers with trimix diluent.

Mike and Alan Fraser are keen anglers and have fished the reef on which the coelacanth was found for many years. They are also avid scuba divers who have been using AP diving rebreathers for the past 9 years and are familiar with the underwater terrain on this coast. Mike and Alan Fraser had previously speculated that the Umzumbe River

Canyon would be an ideal place to spot a coelacanth, because the caves and cracks seen on the sonar would offer good shelter from predators.

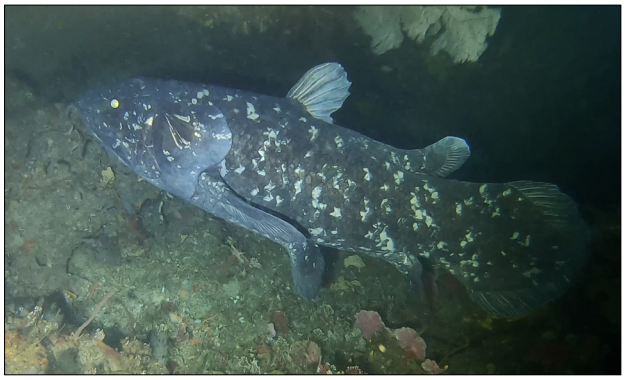


Figure 1: Coelacanth off Pumula on the KwaZulu-Natal South Coast, South Africa, on 22 November 2019 (photo: Bruce Henderson).

The reef on which the dive took place (the longitude and latitude coordinates for the discovery site are known but are being kept confidential in order to safeguard the coelacanth) is about 1 km from the continental shelf edge and is washed by strong currents. The coelacanth was first found by Alan Fraser who was swimming ahead of the other divers. On the video recorded during the dive (supplementary material) he can be heard shouting for Bruce Henderson, who had the GoPro 7 video camera with a 150 m underwater housing. The maximum depth of the dive was 72 m and the total bottom time 15 min, of which about half was spent with the coelacanth. Bruce Henderson filmed the coelacanth at a depth of 69 m.

The single coelacanth that was sighted remained relatively motionless under an overhang despite the attentions of the four divers and their strobe lights. It maintained a head-down position, slowly moving its paired fins. Although the epicaudal ridge along the middle of the tail was prominent, the fin rays on the epicaudal lobe of the tail fin did not extend beyond the curve of the rays on the dorsal and ventral portions of the fin, as in the second coelacanth, which Smith initially thought was a separate species.[11]

The size of the coelacanth was estimated to be 180–200 cm and about 100 kg by comparing its dimensions with those of the divers, although it is difficult to estimate a fish's size accurately under water. It would almost certainly have been a female individual as male coelacanths rarely exceed 150 cm in length.[9,26]

This estimated size is comparable to the largest coelacanths on record which include a 179-cm, 98-kg female caught off Pebane in Mozambique (CCC 162)[14], a 183-cm female caught off Mutsamudu, Anjouan (CCC 126), a 187-cm, 85-kg individual caught off Toliara in Madagascar (CCC 284) and a 190-cm female caught off Chiconi, Anjouan (CCC 127), according to the official CCC inventory[29,30].

What is the significance of the Pumula coelacanth discovery? It indicates that coelacanths live along our coast further south than the iSimangaliso Wetland Park in Maputaland and raises the possibility that they may live elsewhere along the KwaZulu-Natal coast and even further south along the Transkei coast into the Eastern Cape. If this is the case, then the first coelacanth that was caught off East London over 80 years ago may not have been a stray but a member of a resident population. Instead of being washed southwards from the tropics by the south-flowing Mozambique current, as Smith[7] had suggested, coelacanths may have moved purposefully over time into suitable habitats further south than their optimal range in the tropics.

Dives conducted using the research submersible *Jago* off the Eastern Cape coast near East London and Port Elizabeth in 1991 by Hans Fricke, Jurgen Schaüer, Mike Bruton and others revealed that underwater habitats there were suboptimal, with only small overhangs and no deep caves.[9] On these dives large ambush predators, such as the wreckfish, *Polyprion americanus*, seemed to fill the coelacanth's niche.[9] More suitable habitats for coelacanths along the Eastern Cape coast have since been found in the Chalumna Canyon by Dr Kerry Sink of ACEP (2019, personal communication) using remotely operated vehicles near the capture site of the first specimen in December 1938.

The depth preferences of coelacanths throughout their range extends from about 54 m to over 800 m,[21-31] shallow by marine standards as the average depth of the ocean is 3688 m. The Pumula coelacanth record is therefore the second shallowest yet recorded for a healthy, non-pregnant adult coelacanth. Coelacanth specimens, and the data associated with their capture, are well documented; the latest edition of the coelacanth inventory compiled by Rik Nulens for the CCC lists 323 specimens caught to date (Nulens R 2019, personal communication, December 29).

Dead or dying coelacanths, some with their guts full of plastic, one with a *Tetradon* blaasop stuck in its mouth (CCC 182), have been found floating at the water surface off Tanzania.[30] A large coelacanth (179 cm, 98 kg; CCC 162), pregnant with 26 pups, was caught between 40 m and 44 m on the continental shelf off Pebane in northern Mozambique in August 1991.[12,30]

The discovery of a coelacanth at a depth of only 69 m off Pumula (and at 54 m in the Diepgat Canyon) suggests that they may live shallower than previously thought, at least at the southern end of their range, which means that they may be more accessible to mixed-gas divers, as well as to shallow-water remotely operated vehicles and research submersibles, for study. Many aspects of coelacanth biology and behaviour have not been documented, including whether they guard their young after birth and the diet and habitat preferences of the young.

The shallower depths at which coelacanths appear to live off South Africa, compared to the Comoros, Tanzania or Madagascar, may be a consequence of their relatively low tolerance of high water temperatures and low oxygen saturations.[21-26] As oxygen saturations are lower in warmer water, they may tend to live deeper in the warmer waters of the tropics and shallower in the cooler subtemperate waters off South Africa's east coast. Mike Fraser has reported that, although the surface water temperature on the day of the discovery (22 November 2019) was about 25 °C, there was a marked thermocline at about 15 m from the bottom where the recorded temperature fell to 17 °C.

The discovery of a living coelacanth off the South Coast of KwaZulu-Natal also reveals how little we know about marine life off our coast, despite intensive research, increasingly intense deep diving ventures and long-term commercial and recreational fishing pressure. That coelacanths have been living undiscovered off the heavily fished south coast of KwaZulu-Natal, despite the high profile of the fish over the past 80 years, suggests that many remarkable discoveries remain to be made in our oceans.

ACEP and other programmes that promote and facilitate multi-disciplinary research in our marine environment should therefore continue to receive priority financial and logistical support, and recreational divers should be encouraged to collaborate with scientists so that their valuable observations can be included in the scientific dialogue. More effort should also be made to create organised platforms that make it possible for 'citizen scientists', especially deep divers, to participate meaningfully in scientific research. The possibility of creating an offshore Marine Protected Area off Pumula, without unduly impacting on the activities of recreational anglers, also needs to be considered.

Acknowledgements

The dive team thanks Marc Dukes for processing Bruce Henderson's digital video recording, extracting still images therefrom, and producing a video that can be viewed on YouTube by following the link https://www.youtube.com/watch?v=dGP-3MyBwtI

References

1. Bruton MN. Fifty years of coelacanths. S Afr J Sci. 1989;85, 205.
2. Bruton MN. The living coelacanth fifty years later. Trans Roy Soc S Afr. 1989;47:19–28. https://doi.org/10.1080/00359198909520147
3. Bruton MN. The annotated old fourlegs: The updated story of the coelacanth. Cape Town: Struik Nature; 2017.

4. Bruton MN. The Fishy Smiths: A biography of JLB and Margaret Smith. Cape Town: Penguin; 2018.
5. Smith JLB. A living fish of Mesozoic type. Nature. 1939;143:455–456. https://doi.org/10.1038/143455a0
6. Smith JLB. A surviving fish of the order Actinistia. Trans Roy Soc S Afr. 1939;27:47–50. https://doi.org/10.1080/00359193909519786
7. Smith JLB. Old Fourlegs: The story of the coelacanth. London: Longman Green; 1956.
8. Bruton MN. Curator and crusader. The life and work of Marjorie Courtenay-Latimer. Cape Town: Footprint Press; 2019.
9. Bruton MN. When I was a fish: Tales of an ichthyologist. Cape Town: Jacana Media; 2015.
10. Bruton MN. Does the coelacanth occur in the Eastern Cape? Eastern Cape Naturalist. 1989;33(3):5–13.
11. Smith JLB. The second coelacanth. Nature. 1953;171:99–101. https://doi.org/10.1038/171099a0
12. Bruton MN, Cabral AJP, Fricke H. First capture of a coelacanth, *Latimeria chalumnae* (Pisces, Latimeriidae) off Mozambique. S Afr J Sci. 1992;88:225–227.
13. Heemstra PC, Freeman ALJ, Yan Wong H, Hensley DA, Rabesandratana HD. First authentic capture of a coelacanth, *Latimeria chalumnae* (Pisces: Latimeriidae), off Madagascar. S Afr J Sci. 1996;92:150–151.
14. De Vos L, Oyugi D. First capture of a coelacanth, *Latimeria chalumnae* Smith, 1949 (Pisces: Latimeriidae), off Kenya. S Afr J Sci. 2002;98:345–347.
15. Menda A. Coelacanth fish discovered in Tanzania. JUSTA (Journalists Union of Science & Technology Advancement in Africa), 24 September 2003.
16. Benno B, Verheil E, Stapley J, Rumisha C, Ngatunga B, Abdallah A, et al. Coelacanth (*Latimeria chalumnae* Smith, 1939) discoveries in Tanzania. S Afr J Sci. 2006;102:486–490.
17. Anon. Building on the South African Coelacanth Legacy. Grahamstown: South African Institute for Aquatic Biodiversity; 2013.
18. Pouyaud L, Wirjoatmodjo S, Rachmatika I, Tjakrawidjaja A, Hadiaty RK, Hadie W. Une nouvelle espèce de coelacanthe. Preuves génétiques et morphologiques [A new species of coelacanth: Genetic and morphological evidence]. Comptes rendus de l'Académie des Sciences – Sciences de la vie. 1999;322(3):261–267. French. https://doi.org/10.1016/S0764-4469(99)80061-4
19. Erdmann MV. An account of the first living coelacanth known to scientists from Indonesian waters. Env Biol Fish. 1999;54:439–443. https://doi.org/10.1023/A:1007584227315
20. Millot J. First observations on a living coelacanth. Nature. 1955;175:362–363. https://doi.org/10.1038/175362a0
21. Fricke H, Plante R. Habitat requirements of the living coelacanth *Latimeria chalumnae* at Grande Comore, Indian Ocean. Naturwissenschaften. 1988;75:149–151. https://doi.org/10.1007/BF00405310
22. Fricke H, Hissmann K. Natural habitat of coelacanths. Nature. 1990;346:323–324. https://doi.org/10.1038/346323a0
23. Fricke H, Hissmann K. Home range and migrations of the living coelacanth *Latimeria chalumnae*. Mar Biol. 1994;120:171–180. https://doi.org/10.1007/BF00349676
24. Fricke H, Schauer J, Hissmann K, Kasang L, Plante R. Coelacanth *Latimeria chalumnae* aggregates in caves: First observations on their resting habitat and social behavior. Env Biol Fish. 1991;30:281–285. https://doi.org/10.1007/BF02028843
25. Fricke H, Hissmann K, Schauer J, Reinicke O, Kasang L, Plante R. Habitat and population size of the coelacanth *Latimeria chalumnae* at Grande Comore. Env Biol Fish. 1991;32:287–300. https://doi.org/10.1007/BF00007462
26. Balon EK, Bruton MN, Fricke H. A fiftieth anniversary reflection on the living coelacanth, *Latimeria chalumnae*: Some new interpretations of its natural history and conservation status. Env Biol Fish. 1988;23:241–280. https://doi.org/10.1007/BF00005238
27. Ribbink AJ, Robberts MJ. African Coelacanth Ecosystem Programme: An overview of the conference contributions. S Afr J Sci. 2006;102:409–415.
28. Hissmann K, Fricke H, Schauer J, Ribbink AJ, Robberts MJ, Sink K, et al. The South African coelacanths – an account of what is known after three submersible expeditions. S Afr J Sci. 2006;102:491–500.
29. Bruton MN, Coutouvidis SE. An inventory of all known specimens of the coelacanth *Latimeria chalumnae*, with comments on trends in the catches. Env Biol Fish. 1991;32:371–390. https://doi.org/10.1007/BF00007467
30. Nulens R, Scott L, Herbin M. An updated inventory of all known specimens of the coelacanth, *Latimeria* spp. Smithiana Special Publication. 2011;3:1–152.
31. Bruton MN, Stobbs RE. The ecology and conservation of the coelacanth *Latimeria chalumnae*. Env Biol Fish. 1991;32:313–339. https://doi.org/10.1007/BF00007464

AUTHORS:
Mark J. Gibbons[1]
Wayne K. Florence[2]
Maryke Musson[3]
Caragh Barwise[4]
Delphine Thibault[5]

AFFILIATIONS:
[1]Department of Biodiversity and Conservation Biology, University of the Western Cape, Cape Town, South Africa
[2]Department of Research and Exhibitions, Iziko Museums of South Africa, Cape Town, South Africa
[3]Two Oceans Aquarium, Cape Town, South Africa
[4]Craft and Design Institute, Cape Town, South Africa
[5]Institut Méditerranéen d'Océanologie, Aix Marseille Université, Campus de Luminy, Marseille, France

CORRESPONDENCE TO:
Mark Gibbons

EMAIL:
mgibbons@uwc.ac.za

HOW TO CITE:
Gibbons MJ, Florence WK, Musson M, Barwise C, Thibault D. Creating opportunities through science symposia. S Afr J Sci. 2020;116(3/4), Art. #7715, 3 pages. https://doi.org/10.17159/sajs.2020/7715

ARTICLE INCLUDES:
☐ Peer review
☒ Supplementary material

KEYWORDS:
collaboration, conferences, entrepreneurship, public outreach, school outreach

PUBLISHED:
26 March 2020

Creating opportunities through science symposia

For most marine scientists, unless we work in the field of fisheries development or at the interface of science and policy, it is rare to feel that we are making an impact on the lives of people in the wider community. Most research scientists at universities and government laboratories also have limited opportunity to engage with schools and the general public outside of once-a-year open days. But beyond the science and networking, conferences, especially international conferences, can provide a myriad of opportunities for us to redress both these issues in a way that enriches all. Here, we describe the programme of development-related activities that supported the 6th International Jellyfish Blooms Symposium, and their impact, and we urge it be used as a template for other scientific meetings in the future.

Jellyfish are far more than merely an interesting find on the beach. On the one hand, when abundant, jellyfish can cause economic harm to the tourism, aquaculture, fisheries and energy sectors[1], but on the other hand, they provide food for other animals, e.g. turtles, shelter for juvenile fish and a potential resource to exploit[2]. In recognition of their role in marine ecosystems, the international jellyfish community updates and renews itself at a conference every 3 years or so. The first meeting was held in the USA in January 2000 and after conferences in Australia, Argentina, Japan and Spain, Africa's turn came in 2019, after Monty Graham (University of Southern Mississippi) convinced one of us (M.J.G.) to host the conference at the University of the Western Cape.

'Whilst Africa may have witnessed the evolution of *Homo sapiens sapiens*, she is nevertheless young and inexperienced … in so many ways. Consequently, the conference will focus on development in its broadest sense…'. These sentences open paragraph two on the conference webpage (http://www.jellyfishbloom2019.co.za), and effectively encapsulated our thoughts on both the science programme and ancillary activities.

Partnerships

As any who have tried will know, organising a conference single-handedly is impossible. Small universities with a limited staff complement generally have neither the budget nor the skill set to organise conferences, and expertise to advise on non-core activities is missing. The University of the Western Cape is no exception in this regard. Thus, in order to achieve our development goals, the University of the Western Cape actively sought partner institutions with a footprint in public engagement and education outside the tertiary education space: private enterprise (Two Oceans Aquarium) and the local natural history museum (Iziko Museums of South Africa).

Although the Two Oceans Aquarium is a premier tourist attraction at the V&A Waterfront development in Cape Town, it plays a very important role in regional and national education, conservation and research. Together with an active volunteer programme, more than 20 years' experience in adult education and a comprehensive schools programme beginning with the early childhood development phase, it has recently been developing and driving the implementation of a marine sciences curriculum for Grades 10–12, which is already being offered at various pilot schools as an additional subject.

Iziko Museums of South Africa operates 11 national museums in Cape Town, housing natural history, social history and art collections in magnificent historical buildings. The Iziko South African Museum is the oldest and largest museum in South Africa and boasts an impressive history of scholarship and education. There is a strong commitment to engaging with diverse communities through various school outreach and enrichment programmes, as well as empowering and inspiring visitors through education. Half a million people visit Iziko's museums each year and around 50 000 school learners are guided through its galleries by museum educators and curators.

Reaching out

At all previous International Jellyfish Blooms conferences, the only representation from sub-Saharan Africa had been limited to the author (M.J.G.) or his students. But jellyfish are not only a southern African issue[3], and so if, as a community, we are going to develop 'in-house'[4] jellyfish science on the continent, we need to identify and sponsor participation by scientists and students from countries beyond South Africa. Through various contacts at the UN Food and Agriculture Organization, and targeted advertisement, four delegates were identified from elsewhere in Africa: one each from Morocco, Senegal, Ghana and Cameroon. All four are working on jellyfish in their home countries, either as postgraduate students or as part of their work in government laboratories. Sponsorship for their attendance was secured from two sources: the National Research Foundation of South Africa, through a grant to M.J.G., and conference registration fees. It is common practice when organising conferences to fund the costs of plenary speakers using the latter. For this conference, however, we omitted plenary talks, which had the further beneficial effect of making all delegates equal, and used these monies to help cover the costs of African colleagues. Local student mentors were assigned to chaperone and assist each African visitor, for all of whom the trip was the first to South Africa.

After the conference, our African guests participated in a 2-day workshop on jellyfish identification at Iziko's Natural History Museum, led by André Morandini from Saõ Paulo in Brazil and Ilka Straehler-Pohl from Stade-Hagen in Germany. These two experts gave of their time freely to participate, recognising it as an opportunity to build links with Africa for potential future collaborations and for a chance to see recently collected material from the region. This was also an occasion on which to access global experts for taxonomic determination/validation of the jellyfish collections of both the University of the Western Cape and Iziko.

In their feedback, the African delegates had this to say:

> *I have gain*[ed] *a lot both in knowledge and connections. I have the opportunities, now it is the time to go forward … I will use* [the] *contacts I made to carry out collaborative research, applying for grants and fellowships.*
>
> Gisèle FG Youbouni, PhD candidate, Cameroon

The networking I made during the conference was great and the knowledge gained will help improve the quality of my research.

Prisca Adongo Ayerijenna, MPhil candidate, Ghana

Educational outreach

Branding

If you talk to people who work in advertising, they will tell you that branding is everything. While we may be generalising, our experience is that most science students today cannot tell a Monet from a Manet and while they may be able to think creatively in their discussions, their abilities to behave similarly in the arts are generally limited. We turned to the Red and Yellow Creative School of Business for assistance with the conference brand. Red and Yellow is a small tertiary-level institution that trains students, primarily in marketing, advertising and communications. Following an interactive lecture to students on jellyfish and the conference, Red and Yellow gave students the option to submit a design for the conference logo as one of their formal assignments. Almost half the class opted in for the project: staff at the design school selected the shortlist (Supplementary figure 1), from which we identified the winner (Figure 1). All shortlisted designers were invited to the conference dinner, and the winning logo was displayed on all letterheads, bags, T-shirts, etc.

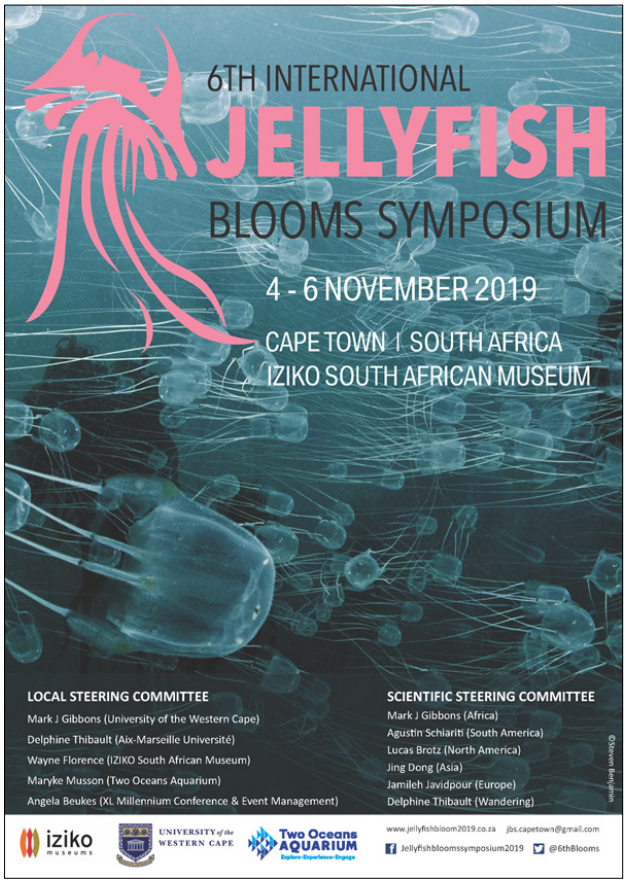


Figure 1: The winning design.

Feedback from educators at Red and Yellow was as follows:

The students enjoyed the experience of working with a real 'client' because they stood a chance of having their work displayed in the world outside the classroom. It's always exciting for students when their work makes it into the professional arena, and they get a taste for their future career.

Connor Cullinan, Red and Yellow

I'm not sure how often these events come around but you are welcome to contact us again and we'll try to help if we can.

Nini van der Walt, Red and Yellow

In designing the conference poster (Figure 1), we approached local wildlife photographer Steve Benjamin. Steve, who is more than your average citizen scientist, happily shared some of his photographs of jellyfish with us and allowed us to use one of his images as a backdrop for the poster.

Public engagement

Open lectures given by selected conference delegates should be mandatory at conferences. At the very least they represent an opportunity for the taxpayer to see how public money is spent; at best they give local audiences a chance to hear about latest developments and to engage with international experts. With the 'Ocean Exhibit' as a backdrop, Two Oceans Aquarium hosted a relaxed evening during which three of the international conference delegates gave 30-min informal talks on: jellyfish venoms, how they work and how to treat them (Angel Yanagihara, Hawaii); interactions between jellyfish and humans (Lucas Brotz, Vancouver); and real and potential uses of jellyfish (Dror Angel, Haifa). The evening was a sell-out!

This was a once-in-a-lifetime opportunity to learn more about jellyfish, friend or foe, from world experts. What a fascinating, informative and fantastic event. Thank you.

Ingrid Sinclair, Two Oceans Aquarium

The conference itself was hosted by Iziko, on the fifth floor of their iconic building at the top of The Company Gardens in central Cape Town. Afternoon poster sessions were held in the whale well, so called because of the whale skeletons hanging from the ceiling. This latter area is open to the public during the day, and day visitors to the museum had the opportunity to engage with the conference indirectly.

Iziko Museums have a group of dedicated educators who work with visiting school groups. Several weeks prior to the conference, during National Marine Week, around 200 school learners were invited to participate in 'JellyArt'; a multidisciplinary activity that saw learners create jellyfish artworks from recycled materials and pen a message/question for the delegates of the symposium to consider. These JellyArt works were displayed during the symposium, further continuing the conversation with the general museum visitors (Supplementary figure 2). During the conference itself, a young-at-heart and charismatic delegate gave up time to present a public lecture entitled 'The Light and Dark Side of Jellyfish' and also participated in school-room classes given by museum educators, enjoying the opportunity to work with a diverse and receptive audience.

Creating commercial opportunities

A few marine biologists are likely to have a tattoo of their study organism somewhere on their body, and many probably have work-themed paraphernalia at home and/or in the office. Recognising this inclination, and given the high levels of unemployment in South Africa, we set out to engage with local entrepreneurs who could provide delegates with locally crafted gifts for retail.

We approached the Craft and Design Institute (CDI), a Cape Town based craft and design sector development agency with a mission to develop capable people and build responsible creative enterprises trading within local and international markets. With funding from the Western Cape Department of Cultural Affairs and Sport, the CDI was able to run a free workshop at the Two Oceans Aquarium, which was well attended by interested crafters, many of whom had never visited the venue before. This day-long event was a themed product design and creativity workshop aimed at stimulating products that could be sourced for the Symposium and followed a short presentation to the attendees on jellyfish with a view to sparking ideas for unique goods.

The result was that seven crafters developed a range of items from earrings, through socks and cushion covers, to lamps, mobiles and children's toys, as well as bags and clothing. Items were available for sale at the poster evenings and all social events: a catalogue of items offered can be found on the conference website. It was the first time that the CDI had been involved in a project such as this and they appreciated the initiative.

Sales were generally good and most crafters indicated that, although there were challenges, it was a success:

> *Such an awesome journey we all took and persevered right to the end … was worth it.*
>
> Lewis Ndlovu, Kingsdale Emporio

> *I'd definitely like to get involved in more conferences around animals and natures!!!*
>
> Nehanda Magan, Nehanda Design

> *I will be happy to do it again with such kind friendly and great people.*
>
> Joseph Kalango, Design Cradle

Feedback from the CDI:

> *This product development opportunity was such a good experience for producers to make money and be successful. I think this can be attributed to the 'narrow', defined parameters of the project. Parameters of a defined theme (that is visually inspirational), a price range, and the security of a definite interested audience. It is an excellent, not too risky a project format for people learning, to be involved in. I want to use this model going forwards as an iterative learning and consolidating experience.*
>
> Fran Stewart, CDI

The way forward

In relatively small conferences such as this (we had 150 delegates from 23 countries), it is customary for prospective hosts of the next meeting to make a bid. In Cape Town, two institutions came forward: the University of British Columbia, Canada and the University of Kerala, India. Immediately prior to voting at the closing dinner, Lucas Brotz announced that the University of British Columbia was withdrawing its offer to host the next meeting, because he felt that the momentum generated around development in Cape Town needed to be carried forward. Although India has a rich jellyfish diversity and harvests almost 100 000 tons annually, there are very few scientists there who study them, primarily owing to a lack of capacity. The striking similarities between India and South Africa suggest that the development focus in Cape Town, could be successfully repeated there, with support from the community.

Support from the community appears to be strong. In our post-symposium questionnaire, we expressly asked whether delegates appreciated the development initiatives we had put together, and the feedback was resoundingly positive. Over 90% of respondents were interested by our engagement with local crafters and 60% purchased at least one item from them; almost 80% of respondents appreciated the school outreach and 57% indicated that they would be happy to participate in such activities at future conferences, time permitting.

Given this level of support from a relatively small international conference, imagine the impact of a much bigger conference. The International Congress of Zoology is meeting in Cape Town in 2020, the 14th International Conference on Copepoda is being held in the Kruger National Park in 2020 and the International Ornithologists Union is gathering in Durban in 2022. Not to mention local conferences, some of which attract more delegates than the specialised international ones. With careful planning, as a socially responsible academy, we can and should provide opportunities for the wider community in as many of our endeavours as possible.

References

1. Purcell JE, Uye S, Lo WT. Anthropogenic causes of jellyfish blooms and their direct consequences for humans: A review. Mar Ecol Prog Ser. 2007;350:153–174.
2. Gibbons MJ, Boero F, Brotz L. We should not assume that fishing jellyfish will solve our jellyfish problem. ICES J Mar Sci. 2016;73(4):1012–1018.
3. Roux J, Van der Lingen CD, Gibbons MJ, Moroff NE, Shannon LJ, Smith AD, et al. Jellyfication of marine ecosystems as a likely consequence of overfishing small pelagic fishes: Lessons from the Benguela. Bull Mar Sci. 2013;89:249–284.
4. Gibbons MJ, Govuza B. The nature of international collaboration in the Benguela upwelling region, 2000-2016. S Afr J Sci. 2019;115(1/2), Art. #5442, 5 pages. https://doi.org/10.17159/sajs.2019/5442

AUTHORS:
Graham I.H. Kerley[1]
Joris P.M.G. Cromsigt[1,2,3]

AFFILIATIONS:
[1]Centre for African Conservation Ecology, Nelson Mandela University, Port Elizabeth, South Africa
[2]Department of Wildlife, Fish and Environmental Studies, Swedish University of Agricultural Sciences, Umeå, Sweden
[3]Copernicus Institute of Sustainable Development, Utrecht University, Utrecht, the Netherlands

CORRESPONDENCE TO:
Graham Kerley

EMAIL:
graham.kerley@mandela.ac.za

HOW TO CITE:
Kerley GIH, Cromsigt JPMG. From reindeer to rhino: Reflections on 'Climate change mitigation and adaptation benefits of wilder rangelands'. S Afr J Sci. 2020;116(3/4), Art. #7604, 2 pages. https://doi.org/10.17159/sajs.2020/7604

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
rangelands, climate smart, mitigation, adaptation, management

PUBLISHED:
26 March 2020

From reindeer to rhino: Reflections on 'Climate change mitigation and adaptation benefits of wilder rangelands'

Ever since the Swedish botanist Anders Sparrman[1] commented in 1786 on the state of overgrazing brought about by sedentary trekboer pastoralists in the southwestern Cape, the optimal use of rangelands in South Africa has been the subject of debate. Two centuries later, Hoffman and Ashwell[2] painted a bleak picture of our stewardship of this precious resource. These rangelands are not merely a source of fibre and food production, but are also key for biodiversity conservation and rural livelihoods. The prospects of global change, and its impacts on these rangelands, are now being explored in the ongoing debate about the sustainable use of rangelands (e.g. Gillson et al.[3]). Part of this discussion has to be on how we can develop new ways of using these rangelands that are both sustainable and of value to society. It is fitting that, in going the full circle of the Swedish connection, a recent workshop was held under the auspices of the 2019 South Africa–Sweden University Forum's (SASUF) Research and Innovation Week that focused on innovative and out-of-the-box ideas around future rangelands.

The workshop, entitled 'Climate change mitigation and adaptation potential of wilder rangelands' was held at the Centre for African Conservation Ecology, Nelson Mandela University and was co-hosted by Joris Cromsigt (Swedish University of Agricultural Sciences) and Graham Kerley (Nelson Mandela University). It brought together researchers, managers and policymakers from South Africa, Botswana, and Sweden with a goal of building networks and encouraging novel debates around rangelands in the light of anthropogenic climate change. Such debate is urgent because rangeland practices are not only increasingly threatened by the changing climate, including issues such as desertification, woody plant encroachment and extreme weather[4], but rangeland practices are also a major driver of anthropogenic climate change[5]. We urgently need novel models for our rangeland management that increase the capacity of rangelands to adapt to changing climates and contribute to mitigating climate change.

Traditionally, discussions on South African rangelands have focused on concepts like grazing management systems, 'carrying capacities' and overgrazing, with a heavy focus on domestic livestock.[6] Equilibrium views of ecosystem dynamics often formed the basis of these earlier discussions, which mostly had a local (regional, biome, national) focus. It is only in the last 30 years that this debate has been extended to include indigenous wildlife[7], disequilibrium dynamics[8], and broader ecosystem services provided by rangelands. The scale has also expanded to consider global issues such as climate change, for example, the role of methane production by domestic herbivores as a climate forcer.[9]

The recent paper by Cromsigt et al.[10] exploring how rangeland management could be integrated into climate change adaptation and mitigation strategies, is a novel but logical development of this debate.

A key lesson from this paper, which served as a catalyst for our workshop, is that the functional diversity, and the spatial-temporal dynamics, of species-rich wild grazing systems offer interesting opportunities for climate change adaptation and mitigation. For example, wild systems include a diversity of non-ruminant species, which emit very little methane. The paper thus concluded that solutions for some climate change challenges might lie in wilder rangelands. Wilder rangelands might entail replacing managed livestock systems with harvested communities of native wild herbivores and/or 'rewilding' livestock grazing practices through learning from wild grazing systems. By extension, management of rangelands may provide avenues to both adapt to climate change as well as to mitigate it.

Workshop participants provided a variety of relevant insights. Joris Cromsigt began with an example of rangelands in Europe's far north where reindeer (*Rangifer tarandus*) may play a role in cooling the tundra because their impacts increase the extent to which these rangelands reflect sunlight (albedo) by turning shrub-dominated tundra into grassland steppe.[11] Importantly, in these regions, the possible cooling effect of shrubs sequestering carbon is offset by the warming effect of their low albedo.[12] Rewilding the Nordic tundra with reindeer may thus be among the more effective ways of mitigating climate change in this region, and is very urgent in the light of the current rapid woody encroachment. Cromsigt then extended the analogy from reindeer to rhino and, based on data from Hluhluwe iMfolozi Park, showed how rhino (and other large wild grazers) may have similar positive effects on the reflective properties of African rangelands: they promote grassy vegetation with high albedo over woody vegetation. Contributions by Kathleen Smart (Rhodes University) and Tony Palmer (Agricultural Research Council) clearly suggested that, similar to the tundra systems, albedo effects may offset the carbon benefits of woody encroachment in South African rangelands, particularly in the arid to semi-arid zone. Shaun Welman (Nelson Mandela University), Marietjie Landman (Sol Plaatie University) and Sheunesu Ruwanza (Rhodes University) highlighted important gaps in our knowledge, such as in terms of grazing–soil interactions and the ecophysiology of wild grazers. Student contributors Tebogo Sebake and Masego Mokobela (Nelson Mandela University) emphasised how a new generation can show the way through their insights and enthusiasm to take on the challenges. There was also a series of talks by local NGOs. Participants also stressed that careful consideration should be given to how and where rewilding of rangelands is done so as to maximise the benefits for rural communities. In addressing this, care needs to be taken that wilder rangelands as a climate change mitigation tool are not imposed on the poorest people for the benefit of the richest, but rather that a new model is provided that increases social equitability and empowerment.

Participation by representatives of government departments with a responsibility towards policy and management of rangelands was a key opportunity for them to provide insights into policy approaches to these issues. It emerged that the approach by the Department of Environmental Affairs (Hlengiwe Mbatha, now Department of Environment, Forestry and Fisheries) to mitigation is heavily invested in carbon sequestration interventions. Suggestions and

insights from this workshop might provide a framework to review critically and broaden this strategy and empower the Department to consider additional opportunities for climate change adaptation and mitigation – in line with the National Climate Change Response policy of tackling climate change mitigation while reducing risks and vulnerabilities. Although not the core focus of the Department of Agriculture, Rural Development and Land Reform (Siviwe Shwababa), including climate change adaptation and mitigation interventions in the land reform programme may provide additional opportunities to manage rangelands sustainably while empowering rural communities. For both these departments, the prospects of linking Green Economy initiatives, and specifically the wildlife aspect, to climate change adaptation and mitigation and sustainable social transformation need to be explored.

The workshop successfully initiated by a small network of scientists and managers in Sweden and South Africa with a shared interest in rangelands and climate change adaptation and mitigation, has refreshed the rangeland debate in South Africa. The gathering concluded that wilder rangelands might provide significant opportunities for climate change mitigation, through carbon storage, ecological restoration, altered fire dynamics, increased albedo and reduced methane production, while achieving equitable social empowerment. Wilder rangelands may also be better adapted to changing climatic conditions, and as such, would not be yet another ecological restoration model separating humans from the natural environment, but instead might investigate the potential of wilder landscapes to include people. Novel ways of using our rangelands should be co-created through close interaction among scientists, policymakers and local communities. Critically, however, what is needed for policy and management interventions to be robust, are evidence-based insights into how wilder rangeland management practices can influence the drivers of climate change to the benefit of society, while at the same time providing adaptation responses to deal with ongoing climate change. As a next step, arising from this workshop, we will continue this initiative through the SASUF mechanism, by convening a meeting in Sweden in 2020. Clearly, this broader and ambitious opportunity needs to be developed further.

Acknowledgements

We thank SASUF for funding the workshop, and Mariska te Beest, Kent Buchanan, Christo Fabricius, Cleo Graf, Marietjie Landman, Tony Palmer, Mzwamadoda Mbangwa, Hlengiwe Mbatha, Masego Mokobela, Manyano Mpelame, Sheunesu Ruwanza, Tebogo Sebake, Siviwe Shwababa, Kathleen Smart and Shaun Welman for their enthusiastic and insightful contributions to this workshop. This work emerges from the 'Wilder Rangelands for a Climate-Smarter Future' research programme led by Mariska te Beest, Joris Cromsigt and Graham Kerley, a collaboration between Nelson Mandela University and the Swedish University of Agricultural Sciences.

References

1. Sparrman A. A voyage to the Cape of Good Hope, towards the Antarctic polar circle, and round the world; but chiefly into the country of the Hottentots and Caffres, from the year 1772 to 1776. Vols 1 and 2. 2nd edns. London: Robinson; 1786.
2. Hoffman MT, Ashwell A. Nature divided: Land degradation in South Africa. Cape Town: UCT Press, 2001.
3. Gillson L, Midgley GF, Wakeling JL. Exploring the significance of land-cover change in South Africa. S Afr J Sci. 2012;108(5–6), Art. #1247, 3 pages. https://doi.org/10.4102/sajs.v108i5/6.1247
4. Asner GP, Elmore AJ, Olander L P, Martin RE, Harris AT. Grazing systems, ecosystem responses, and global change. Ann Rev Environ Resour. 2004;29:261–299. https://doi.org/10.1146/annurev.energy.29.062403.102142
5. Arneth A, et al. Climate change and land: An IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems: Summary for policymakers. New York: GMO & UNEP; 2019.
6. Tainton NM, editor. Veld and pasture management in South Africa. Pietermaritzburg: Shuter & Shooter/Natal University Press; 1981.
7. Carruthers J. 'Wilding the farm or farming the wild'? The evolution of scientific game ranching in South Africa from the 1960s to the present. Trans Roy Soc S Afr. 2008;63:160–181. https://doi.org/10.1080/00359190809519220
8. Hawkins H-J, Short A, Kirkman KP. Does Holistic Planned Grazing™ work on native rangelands? Afr J Range Forage Sci. 2017;34:59–63. https://doi.org/10.2989/10220119.2017.1367328
9. 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;43:247–254. https://doi.org/10.4314/sajas.v43i3.2
10. Cromsigt JPGM, Te Beest M, Kerley GIH, Landman M, Le Roux E, Smith FA. Trophic rewilding as a climate change mitigation strategy? Phil Trans R Soc B. 2018;373, Art. #20170440, 12 pages. http://dx.doi.org/10.1098/rstb.2017.0440
11. Te Beest M, Sitters J, Ménard CB, Olofsson J. Reindeer grazing increases summer albedo by reducing shrub abundance in Arctic tundra. Environ Res Lett. 2016;11, Art. #125013, 13 pages. https://doi.org/10.1088/1748-9326/aa5128
12. De Wit HA, Bryn A, Hofgaard A, Karstensen J, Kvalevåg MM, Peters GP. Climate warming feedback from mountain birch forest expansion: Reduced albedo dominates carbon uptake. Glob Change Biol. 2014;20;2344–2355. https://doi.org/10.1111/gcb.12483

AUTHOR:
Shari Daya[1]

AFFILIATION:
[1]Environmental and Geography Science, University of Cape Town, Cape Town, South Africa

CORRESPONDENCE TO:
Shari Daya

EMAIL:
shari.daya@uct.ac.za

HOW TO CITE:
Daya S. On shaky ground: A response to Long et al. (2019). S Afr J Sci. 2020;116(3/4), Art. #7932, 3 pages. https://doi.org/10.17159/sajs.2020/7932

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
Geography, decoloniality, Africanisation, higher education

PUBLISHED:
26 March 2020

On shaky ground: A response to Long et al. (2019)

Geography's histories are indubitably imperialist. The very origins of the discipline, the 'writing of the world' that its name denotes, lie in colonial projects of cartography, exploration, knowledge and conquest.[1] The contexts in which we in contemporary South Africa practise geographical scholarship remain oppressive, as the recent Fallist movements have demonstrated and as gamEdze and gamedZe[2], for example, clearly show. There is an urgent need in South African higher education for decolonised and transformed spaces of teaching, learning and research. It was therefore with interest that I read Long et al.'s[3] call for the decolonisation of Geography curricula in South African universities. I found myself ultimately frustrated, however, by the lack of clarity in their Commentary about what decolonisation might mean for our discipline, as well as concerned by the authors' misreadings of postcolonial theory and their insular and exclusionary ideas of 'legitimacy' in academic spaces. In this rejoinder, I focus on what I consider to be some serious flaws in their argument, followed by a brief proposal for more productive decolonisation strategies in our context.

It is worth noting that what we call Human Geography, although a relatively young sub-discipline, has for some time taken the violence of colonialism seriously. As early as 1938, Carl Sauer[4] examined the 'destructive exploitation in modern colonial expansion' and since the 1980s, critical accounts of 'Geography's empire'[5] and its implications for contemporary scholarship have animated debate in journals, books and conferences. Nonetheless, it is true that questions of colonial oppression and its complex legacies have been less prominent in South African geographical scholarship until recently. The silences here, and the slow demographic transformation of South African Geography departments clearly demonstrate the need to render more visible the oppressions and exclusions that shape the foundations of our discipline.

This is the need that Long et al.[3] identify, calling specifically for explicit recognition in curricula of 'local' knowledges and place names, and for a turn to postcolonial theory as 'resistance to imperialism, a critique of the colonial, and a rediscovery of indigenous histories and heritage' (p.1). There would surely be little argument against these, at least as principles, in any contemporary South African department of Geography, and indeed Long et al.[3] would find little evidence in academic writing for their straw man that the postcolonial is 'often misconceived' as the era after colonialism (p.1). However, simply calling for resistance, critique and rediscovery does not amount to a substantive call for decolonised curricula (or indeed research).

More problematically, however, the article draws 'postcolonial theory' into service as a shorthand for both 'decolonisation' and 'Africanisation', without making it clear how the authors understand any of these terms. I cannot in this short riposte elaborate on the debates about decolonisation versus Africanisation (see, however, Mbembe[6] and Long[7]), but the conceptual slippage in Long et al.'s[3] article undermines their core argument. The authors would do well to take into account Fanon's[8] critique of Africanisation as an ideology that becomes 'more and more tinged by racism' (p. 154) and often leads to what we in South Africa would now term xenophobic violence against fellow Africans. As I point out later, such discursive violence is a real danger in this article.

In the space available here, however, I want to focus less on the 'what' and more on the 'how' in Long et al.[3] – in other words on what these authors see as the *means* by which decolonised curricula should be achieved in South African Geography departments. It is here that their argument steps onto shaky intellectual ground, displaying a surprising ignorance about the scholarship the authors claim as foundational to their project, and putting forth, frankly, regressive proposals for transformation.

Given the centrality of postcolonial theory to the argument, it is reasonable to expect references to theorists such as Said[9], Spivak[10] and Bhabha[11], whose work in cultural and literary studies developed the foundational theories of postcolonialism. The work of the 'Subaltern Studies' scholars, including Chakrabarty[12] and Prakash[13], as well as contemporary postcolonial geographies such as those elaborated by Raghuram, Madge and Noxolo[14-17], Sidaway[18,19], McEwan[20,21], Jacobs[22] and Blunt and Rose[23], to name only a few, similarly escape mention by Long et al.[3] Instead, the authors devote about 10% of their contribution to Carter's[24] critical examination of the 'unwitting and lingering colonial referents' (p. 825) within science education scholarship on cultural diversity. Carter's piece makes a valuable contribution to the specific project of interrogating the language used in science education, but her work cannot substitute for a full appreciation of the wealth of debate within postcolonial geographies nor, more worryingly, is it accurately represented by Long et al.[3] These authors misrepresent Carter's careful work on difference by using her framework as a basis for the claims they go on to make about who should speak within South African higher education, a point to which I return shortly.

While Long et al.[3] do not specify what they mean by decolonisation, they are somewhat clearer about how they think it should be achieved. Some suggestions, such as re-thinking 'criteria of excellence' and building African partnerships, indicate a reflective and inclusive approach. Worryingly, however, they also include proposals to 'question the legitimacy of white geographers to speak on issues of decoloniality', 'leaving them as outsiders to the debates on decolonialisation', and the injunction to stop 'importing…teaching staff and qualifications from overseas' (p. 2).

Again, these propositions merit more reflection than I can give here, but my main critique of Long et al.'s vision of how decolonisation should be effected is that their ideas of 'legitimacy' (a term used six times in the paper) rely heavily on superficial notions of identity and difference. Long et al. employ (1) essentialist ideas of whiteness and indigeneity, and (2) assumptions about which universities house 'students and academics who are truly able to be the African voice that can speak back to the colonial curriculum', presumably based on their race and class identities. By contrast, Carter[24], on whom their work relies so heavily, explicitly points out that postcolonial theory's 'framework of hybridity and ambivalence' gives us 'more complex conceptualizations of cultural difference as hybridized and fluid, always in the making'. Instead of the 'homogenized identities' that Carter explicitly critiques

and that Long et al. in fact perpetuate, we see in Carter's work, 'multiple, mobile, and provisional constructions' (p. 833).

As Carter[24] (p. 832) points out: 'postcolonial interpretation would reveal Western and non-Western borders to be profoundly ambivalent constructs'. Her conceptualisation of difference is in keeping with those of Said, Spivak, Bhabha and the Subaltern School who, for all their theoretical differences, would agree that a postcolonial project must recognise the ways in which, as Gregory[25] put it, colonised and colonial societies always were and remain entangled in 'webs of affinity, influence and dependence' (p. 614). Long et al., by contrast, seem to read difference as binary, shaped by racial and national/regional boundaries as well as by simplified institutional histories. In such a conceptualisation of difference, it may seem possible to step outside of the histories of colonialism and apartheid to occupy an unsullied critical position. But as postcolonial theorists continually remind us, postcolonial projects can only work '*in medias res*'[26] or 'in the midst of things'. That is, our scholarship can only proceed from inside an ongoing narrative and from within national and cultural spaces that have been thoroughly intertwined for centuries.

It could be argued that the flaws in this article are mostly innocuous, or that their effects are confined to the ivory tower. The oddly gendered language in references such as 'the West and her canon' (p. 1) may not matter much. Perhaps the lack of evidence for the authors' claims about existing curricula in South African Geography is not vital within their larger argument. Even the theoretical lacunae evident in this piece and its misrepresentation of Carter's work could be dismissed as not having material consequences. More problematic, however, is a tendency towards the exclusionary construction of the foreigner that Fanon[8] warns about in his critique of Africanisation. In South Africa's contemporary climate of xenophobia and intense and pervasive violence,[27-30] lazy references to 'importing' people from 'overseas' and curricula 'imported… from foreign nations' are politically irresponsible (see Naicker[31] and Tagwirei[32] for analyses of the language of 'foreignness' in South Africa). The clear boundaries that Long et al.[3] draw between Africa and the West, and who may and may not 'legitimately' speak in intellectual spaces, are not only far from the postcolonial theoretical tradition, they are in themselves discursive acts of violence that perpetuate simplistic and damaging notions of identity, difference and belonging.

I hope that this rejoinder to Long et al. will not be misread. I fully support their call for the critical dismantling of colonial iconography, canons, curricula, pedagogies and intellectual paradigms, just as I wish to see more African and South–South partnerships, greater recognition of diverse forms of knowledge, the development of new criteria for excellence in theory, methodology and pedagogy, and the demographic transformation of the staff in South African higher educational institutions. I am firmly of the view that postcolonial theory has much to offer decolonisation projects in South African Geography, because it has already invigorated a wealth of research that is sensitive to 'the colonial present'.[33] The power of that theory lies precisely in the nuanced understandings of historical and experiential difference[34] that emerge from it, rather than its endorsement of Long et al.'s[3] seemingly clear boundaries between 'within and 'outside' (p. 2), or 'local' and 'Western'. No critical scholar should be endorsing ideas of who may 'legitimately' speak in academic spaces that are based on essentialist and reductive notions of 'race' or geographical location.

How then might we proceed in a more productive way to decolonise Geography? I offer three brief suggestions. First, we need to build an evidence base of curricula and pedagogical practices in the discipline, in and beyond South Africa. Knight's[35] recent work – which Long et al. do not cite – is a good start, and although his 'detailed case study' is not as broad as we might wish, it provides a framework for a wider scoping study of Geography teaching.

Second, Geography departments must critically interrogate and re-imagine our pedagogies, as a collective project. Conventional models of teaching in raked lecture halls, and our default modes of assessment (tests, essays, exams) shape our hierarchical and often silencing institutional culture, the revision of which forms a core part of decolonisation.[35] Of course, the effects of such teaching are not always exclusionary, and equally many departments already undertake fieldtrips, workshops, and other forms of teaching that do not rely on a 'container' model of education. *Sharing* these techniques and experiences as part of a collective conversation, in order to build more equitable institutions, is essential to pedagogical learning within our discipline.

A final strategy for decolonisation, one that falls squarely within the ethos of postcolonial theory, is critically to historicise and contextualise the discipline from the perspective of Africa and the Global South. The biographies of power that shape both Physical and Human Geographies, as they are researched and taught in our institutions, remain under-examined. As Mufti[36] argues, we need to interrogate the 'culture of theory' with which we work, not only in our research but also in our teaching.

The call that Long et al.[3] are making, and the thoughts I have sketched out here, merit much more reflection and action. They could form part of an inclusive conversation within our discipline that I, and I trust many others, would embrace.

References

1. Clayton D. Imperial geographies. In: Duncan JS, Johnson NC, Stein RH, editors. A companion to cultural geography. Malden, MA: Blackwell; 2004. p. 449–468.
2. gamEdze and gamedZe. Anxiety, afropessimism, and the university shutdown. S Atlantic Q. 2019;118:215–225. https://doi.org/10.1215/00382876-7281732
3. Long D, Dalu MS, Lembani RL, Gunter A. Shifting sands: The decoloniality of geography and its curriculum in South Africa. S Afr J Sci. 2019;115(9/10), Art. #5738, 3 pages. https://doi.org/10.17159/sajs.2019/5738
4. Sauer C. Destructive exploitation in modern colonial expansion. In: Compte Rendus du Congrès International de Géographie, Amsterdam, 1938. Vol. 2. Leiden: Brill; 1938. p. 494–499.
5. Driver F. Geography's empire: Histories of geographical knowledge. Environ Plann D Soc Space. 1992;10:23–40. https://doi.org/10.1068/d100023
6. Mbembe AJ. Decolonizing the university: New directions. Arts Humanities Higher Educ. 2016;15(1):29–45. https://doi.org/10.1177/1474022215618513
7. Long W. On the Africanization of psychology. S Afr J Psychol. 2016;46(4):429–431. https://doi.org/10.1177/0081246316650853
8. Fanon F. The wretched of the earth. New York: Grove Weidenfeld; 1963.
9. Said E. Orientalism. New York: Random House; 1978.
10. Spivak G. A critique of postcolonial reason: Toward a history of the vanishing present. Cambridge, MA: Harvard University Press; 1999. https://doi.org/10.2307/j.ctvjsf541
11. Bhabha H. The location of culture. New York: Routledge; 1994.
12. Chakrabarty D. Provincializing Europe: Postcolonial thought and historical difference. Princeton, NJ: Princeton University Press; 2000.
13. Prakash G. Subaltern studies as postcolonial criticism. Am Hist Rev. 1994;99:1475–1490. https://doi.org/10.2307/2168385
14. Noxolo P. Postcolonial approaches to development. In: Grugel J, Hammett D, editors. The Palgrave handbook of international development. London: Palgrave Macmillan; 2016. p. 41–53. https://doi.org/10.1057/978-1-137-42724-3_3
15. Raghuram P, Madge C. Towards a method for postcolonial development geography: Possibilities and challenges. Singapore J Trop Geogr. 2006;27:270–288. https://doi.org/10.1111/j.1467-9493.2006.00262.x
16. Raghuram P, Madge C, Noxolo P. Rethinking responsibility and care for a postcolonial world. Geoforum. 2009;40(1):5–13. https://doi.org/10.1016/j.geoforum.2008.07.007
17. Raghuram P, Noxolo P, Madge C. Rising Asia and postcolonial geography. Singapore J Trop Geogr. 2014;35:119–135. https://doi.org/10.1111/sjtg.12045
18. Sidaway J. Postcolonial geographies: An exploratory essay. Prog Human Geogr. 2000;24(4):591–612. https://doi.org/10.1191/030913200100189120
19. Sidaway J, Woon CY, Jacobs J. Planetary postcolonialism. Singapore J Trop Geogr. 2014;35(1):4–21. https://doi.org/10.1111/sjtg.12049

20. McEwan C. Cutting power lines within the palace? Countering paternity and eurocentrism in the 'geographical tradition'. Trans Inst Br Geogr. 1998;23:371–384. https://doi.org/10.1111/j.0020-2754.1998.00371.x

21. McEwan C. Gender, geography and empire: Victorian women travellers in West Africa. Aldershot: Ashgate; 2000.

22. Jacobs J. After empire? In: Anderson K, Domosh M, Pile S, Thrift N, editors. The handbook of cultural geography. London: Sage; 2003. p. 346–353.

23. Blunt A, Rose G, editors. Writing women and space: Colonial and postcolonial geographies. London: Guildford; 1994.

24. Carter L. Thinking differently about cultural diversity: Using postcolonial theory to (re)read science education. Sci Educ. 2004;88(6):819–836. https://doi.org/10.1002/sce.20000

25. Gregory D. Postcolonialism. In: Johnston RJ, Gregory D, Pratt G, Watts M, editors. The dictionary of human geography. 4th ed. Malden, MA: Blackwell; 2000. p. 612–615.

26. Prakash G. Another reason: Science and the imagination of modern India. Princeton, NJ: Princeton University Press; 1999.

27. Crush J, Chikanda A, Skinner C. Mean streets: Migration, xenophobia and informality in South Africa. Waterloo, Ontario: Southern African Migration Programme; 2015. https://doi.org/10.2307/j.ctvh8r45r

28. Dodson B. Locating xenophobia: Debate, discourse, and everyday experience in Cape Town, South Africa. Afr Today. 2010;56(3):2–22. https://doi.org/10.2979/aft.2010.56.3.2

29. Kihato CW. Migrant women of Johannesburg: Everyday life in an in-between city. New York: Palgrave Macmillan; 2013. https://doi.org/10.1057/9781137299970

30. Steinberg J. Xenophobia and collective violence in South Africa: A note of scepticism about the scapegoat. Afr Stud Rev. 2018;61(3):19–134. https://doi.org/10.1017/asr.2018.56

31. Naicker C. The languages of xenophobia in post-apartheid South Africa: Reviewing migrancy, foreignness, and solidarity. Agenda. 2016;30(2):46–60. https://doi.org/10.1080/10130950.2016.1215647

32. Tagwirei C. State narratives of 'foreignness' and 'criminality' in South Africa. Communicatio. 2016;42(2):191–209. https://doi.org/10.1080/02500167.2016.1167751

33. Gregory D. The colonial present: Afghanistan, Palestine, Iraq. Malden, MA: Blackwell; 2004.

34. Roy A. Who's afraid of postcolonial theory? Int J Urban Regional Res. 2016;40(1):200–209. https://doi.org/10.1111/1468-2427.12274

35. Knight J. Decolonizing and transforming the Geography undergraduate curriculum in South Africa. S Afr Geogr J. 2018;100(3):271–290. https://doi.org/10.1080/03736245.2018.1449009

36. Mufti A. Global comparativism. Crit Inquiry. 2005;31(2):472–489. https://doi.org/10.1086/430976

AUTHORS:
Dianne Long[1]
Mwazvita T.B. Dalu[2,3]
Reuben L. Lembani[2,3]
Ashley Gunter[2]

AFFILIATIONS:
[1]School of Education, Social and Economic Science Division, University of the Witwatersrand, Johannesburg, South Africa
[2]Department of Geography, College of Agriculture and Environmental Science, University of South Africa, Pretoria, South Africa
[3]National Research Foundation, Pretoria, South Africa

CORRESPONDENCE TO:
Dianne Long

EMAIL:
dianne.long@wits.ac.za

HOW TO CITE:
Long D, Dalu MTB, Lembani RL, Gunter A. Responding to 'On shaky ground', Daya (2020). S Afr J Sci. 2020;116(3/4), Art. #7930, 1 page. https://doi.org/10.17159/sajs.2020/7930

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
Geography, decolonisation, post-colonial theory

PUBLISHED:
26 March 2020

Responding to 'On shaky ground', Daya (2020)

It was with eagerness that we received a response to our original opinion piece entitled 'Shifting sands: The decolonality of geography and its curriculum in South Africa'.[1] This was composed as a way to engage in the debate around geography, its history, and where it currently stands in relation to student calls for its decolonisation. With this debate in mind, we South African geographers are making progress in moving towards a curriculum that is being scrutinised and, with wilful effort, being made to move forward.

In this reply we seek to address issues we feel are the essence of Daya's[2] response. In doing so, we wish to re-orientate ourselves to the original spirit in which our opinion piece was written. In refocusing our article on its original intent, we have placed most of our focus on addressing three main concerns raised by Daya[2], namely: (1) a lack of clarity on what decolonisation might mean for our discipline; (2) misreadings of post-colonial theory; and (3) insular and exclusionary ideas of legitimacy in academic spaces; with the latter two being our main focus and addressed together.

As authors it was not our intention to incite violence or to create a system of legitimacy that leads to exclusion. The core of our argument was that this has already been done. Much to our dismay, Daya[2] misconstrues our intent to question positionality, and indeed questioning the power of negotiation in debates within decolonisation in geography and in postcolonial theory of different races. Daya[2] inflates the idea of 'othering' and exclusion that she introduces through Africanism, by asserting (rather misleadingly) that the legitimacy of geographers who are white rather than black, 'is questioned'. The question is more about language, and whose voice is likely to dominate, given the current structures within the discipline which are currently still skewed according to pre-apartheid biases, having been influenced by the historical traits of the education system. As Young, in Rukundwa and van Aarde[3], reminds us, 'the language of postcolonial theory is uncompromising, because it threatens privileges and power'.

The main point is whether we should continue to engage in this debate, whilst still completely disregarding that we have yet to address the issues of power within our academic disciplines. Our opinion was that each academic in the discipline of Geography needs to be aware of their power and position in the academy, acknowledge where it comes from and acknowledge its possible limitations for decolonisation. That power then should be utilised not only to engage with ideas around decolonisation but also to engage with scholars and students who are not afforded the same academic recognition in order to promote African research and researchers. Recognition can be limited by one's position in terms of university rankings and the ability to attract funding. This can result in the negligible recognition of African academics at lower ranking institutions and their ideas, which go unread or unpublished.[4] Collaboration could be used as a tool to address this and to promote African research and researchers. At the heart of our original piece was a call to collaboration – an appeal to bring together African scholars, students and ideas that may go unpublished or unread and bring them into the African geography canon to be recognised, taught and researched.

Denying power, or one's ability to do something, on the basis of 'superficial notions of identity and difference' does not assist in collaborative effort to decolonise the Geography curriculum. In the same way, excluding certain academics from the conversation of decolonisation owing to 'superficial notions of identity and difference' also hinders attempts at progress. It was in this spirit that the original Commentary sought to honestly reflect on the race and power dynamics in academic geography and put out a call for collaboration that seeks to address this imbalance.

Daya[2] argues that 'simply calling for resistance, critique and rediscovery does not amount to a substantive call for decolonised curricula (or indeed research)'. However, in response to the article as a whole, Daya[2] proposes three recommendations that are 'more productive'. In fact, these three recommendations mirror very closely those we made in the original opinion piece. We are pleased, therefore, to note that there is some consensus on the possible way forward for decolonising the geography canon in South Africa. However, Daya[2] criticises the means by which we have arrived at these similar recommendations. Moreover, she points out that we have not engaged with the wealth of writings by postcolonial scholars. We feel, however, that the majority of her critique centres on the theory of Africanism she introduces in her article, rather than on postcolonial theory itself that was central to our original article.

We acknowledge that our Commentary did not extensively engage with postcolonial theory, but it was not our intention to do so. Our aim was to plot a potential way forward and seek to interrogate the methodology we proposed. We further aimed to stimulate conversations with other academics in Africa. A rigorous theoretical paper would be a welcome collaboration with anyone who wishes to engage.

Therefore, in the same spirit in which we wrote our article, we welcome conversation, collaboration and bringing together African researchers, educators and scholars to further this African debate and to flesh out a unified way forward to achieve what we can debate and define ourselves as decolonised and Africanised geographers.

References

1. Long D, Dalu MS, Lembani RL, Gunter A. Shifting sands: The decoloniality of geography and its curriculum in South Africa. S Afr J Sci. 2019;115(9/10), Art. #5738, 3 pages. https://doi.org/10.17159/sajs.2019/5738
2. Daya S. On shaky ground: A response to Long et al. (2019). S Afr J Sci. 2020;116(3/4), Art. #7932, 3 pages. https://doi.org/10.17159/sajs.2020/7932
3. Rukundwa LS, Van Aarde AG. The formation of postcolonial theory. Herv Teol Stud. 2007;63(3):1171–1194. https://doi.org/10.4102/hts.v63i3.237
4. Dalu M, Gunter A, Wasserman R, Dalu T. Owning the lake, not just the rod: The continuing challenge of 'the old boys' in knowledge production. S Afr J Sci. 2018;114(7/8), Art. #a0275, 2 pages. https://doi.org/10.17159/sajs.2018/a0275

AUTHORS:
Anita Bosch[1]
Shimon Barit[1]

AFFILIATION:
[1]University of Stellenbosch Business School, Cape Town, South Africa

CORRESPONDENCE TO:
Anita Bosch

EMAIL:
abosch@sun.ac.za

DATES:
Received: 07 Aug. 2019
Revised: 16 Jan. 2020
Accepted: 17 Jan. 2020
Published: 26 Mar. 2020

HOW TO CITE:
Bosch A, Barit S. Gender pay transparency mechanisms: Future directions for South Africa. S Afr J Sci. 2020;116(3/4), Art. #6772, 6 pages. https://doi.org/10.17159/sajs.2020/6772

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:
women's pay, gender wage gap, gender policy

FUNDING:
University of Stellenbosch Business School

Gender pay transparency mechanisms: Future directions for South Africa

The gender pay gap – that is, the difference in wages between men and women for the same or substantially the same work, or work of equal value – still features prominently as a stumbling block in achieving South African gender equality. If South Africa is to dislodge its stagnant gender pay gap, especially for women in the middle and upper levels of the wage distribution, pay transparency – making gender differences in wages known to employees, government and the public – can compel employers to remunerate fairly and equally. We undertook a comparison between the global and national mechanisms of gender pay transparency to propose a way forward to increase transparency in gender pay for South Africa. In addition to a discussion of existing mechanisms, a summary of the gender pay transparency mechanisms of 16 countries is provided as supplementary material to the article. We found that South Africa could strengthen legislated transparency mechanisms, especially with regard to pay reporting and pay audits, provided that sanctions are attached to non-delivery of these duties. Reigniting the debate on strengthening and improving South African legislation and interpretation of existing governance codes in relation to the implementation, monitoring and enforcement of gender pay transparency mechanisms could strengthen the existing collective bargaining framework and provide the impetus to demonstrate that South Africa sees gender equality as an achievable reality, not an improbable ideology.

Significance:

- Despite the presence of constitutional rights and enabling legislation to prevent workplace gender discrimination, South Africa continues to see a stagnant median gender pay gap of between 23% and 35%.
- This study provides a global analysis to reinvigorate the debate on how South Africa can strengthen transparency mechanisms to close the gender pay gap.
- Legislators, activists, board members, trade unions, academics, and organisational leaders are provided with suggestions on transparency mechanisms to improve efforts towards economic gender equality for South Africa.

Introduction

We consider how transparent mandatory gender pay or wage gap mechanisms, as implemented around the globe, could inform future directions in closing the gap in South Africa. The gender pay gap relates to differences in wage earnings between women and men, and is evident at two levels: (1) job-specific gaps (or job and occupational families), in which equal pay for the same or similar jobs, or jobs of equal value is determined, and (2) at the aggregated level, where organisational, industry/sector or national data, over time, show wage gap patterns between women and men.

A recent study by Mosomi[1], utilising 23 years' South African aggregated labour force panel data over a period ending 2015, confirms that the median gender pay gap was stubbornly stagnant in this period after controlling for level of education, amongst other relevant characteristics. More recently, the International Labour Organization (ILO) reported that, in 2017, the median gap was 28.8% based on hourly wages, and 30.3% based on monthly earnings.[2] These statistics show South Africa to be performing poorly in addressing the gap, given that the average global gap is approximately 20%.[2] Mosomi's study also provides varied results for labour force participation at, especially, the lower and upper ends of the wage distribution. At the 10th percentile, the gap has narrowed, which may be a result of collective bargaining[3] and the introduction of national minimum wages[1] for those in the lowest-paying and vulnerable jobs. At the 90th percentile, the wage gap for women at the executive level of organisations initially seemed to be closing, but, in the last 10 years of the Mosomi data set, the gender wage gap widened, and is continually increasing for these women. With the exception of employers that utilise standardised wages, such as in the public sector, where a small but definite gap remains, the South African gender pay gap continues to be problematic.

Whilst South Africa could be regarded as a champion of gender equality in Africa (with a first-place ranking on the Africa Gender Equality Index[4] and the 19th place globally amongst 149 countries on the Gender Gap Index[5]), it is in the details that the picture regarding South African women's economic empowerment is troubling. When turning to household composition, we start understanding why South Africa's gender wage gap as a factor in gender equality issues is especially deserving of attention.

It is estimated that 37.9% of South African households are headed by women[6], female-headed households are approximately 40% poorer than those headed by men[7], and 48.2% of female-headed households support extended family members, in comparison to 23.1% of male-headed households doing the same[8]. Women prioritise spending on household and parenting responsibilities, and they have a longer life expectancy than men.[9] Because women in South Africa support many children and extended family members, and are more likely than men to be employed in occupations with the lowest wages[10], 'finding a man and then sticking to him, is often as much a matter of economic necessity as it is a romantic choice'[11]. It is in economic dependency that power relations are skewed in favour of men, where domestic violence, which is alarmingly high in South Africa[12], may occur and be perpetuated. Beyond household structure, the shaping of women's identities regarding their societal worth and contribution, as

well as their access to resources to improve their own and others' lives, makes the attainment of pay parity significant.

In South Africa, the private sector labour market is largely market-driven.[13] When there is greater reliance on the market, the impact of pay distortions may be increased, as the inherent role of transparent criteria, as enforced through regulatory rules, is reduced, instead favouring subjective criteria.[13]

Wage transparency constitutes what is known to others[14] about wages, beyond what is known to the employer. Given that 'policies and measures to reduce gender pay gaps…need constant monitoring and reinforcement'[15], we set out to do a comparison between the global and national mechanisms of gender pay transparency, to propose a way forward to increase transparency in gender pay for South Africa.

Method

Data sampling, collection and analysis

This study was based on publicly available data on the Internet in English. Data constitute digital literature found on webpages and in reports and documents collected from websites of legislative and governance bodies, utilising the keywords 'gender pay gap public reporting' and 'gender pay gap transparency'. Two main documents provided the springboard for specific country research: the 2018 DLA Piper Report[16] and the 2017 European Commission Report[17]. The 16 countries (excluding South Africa) selected to form part of this study had some form of legislated gender wage transparency obligation allocated to employers operating in those countries. The countries are: Canada (federal), Australia (Commonwealth), Scandinavian countries, the United Kingdom (UK), and countries identified by the European Commission (EC) of the European Union (EU) as having implemented a form of gender wage gap reporting in terms of its 2014 Recommendation.[18] Although 'great accuracy cannot compensate for inaccessibility'[19], we attempted to mitigate the possibility of missing important country data by doing an additional Internet keyword search using the words "gender pay gap" OR "gender wage gap" AND the name of each of the 192 countries that are recognised by the United Nations as member states[20]. Country-level data are provided in Supplementary tables 1–3. India, Peru and Japan were added to the country data as Supplementary table 3, as they have initiated equality reporting measures. The United States of America does not feature in this comparison, because, despite plans to introduce a reporting mechanism, the Trump presidency has put these plans on the back burner.[16] In addition, we consulted library databases ISAP Sabinet, Jutastat and SA ePublications to supplement our Internet searches.

Ethics and permission

'Manual, nonautomated access of information on publicly available web pages should be acceptable without special permissions or actions'[21(p.607)]. Ethical approval to conduct the study was sought, and the study was exempted by the University of Stellenbosch Business School (USB-2019-10369).

Results

Countries in South America (other than Peru), Africa (other than South Africa), and Asia (other than India and Japan) do not have mandatory reporting obligations[22] – which is most likely a reflection of various aspects relating to development, including under-developed formal labour markets and a poor track record of gender rights[23,24].

Process of implementation

The implementation of the reporting mechanisms has been found to take one of two forms: (1) blanket implementation by all qualifying companies, regardless of size, from the date imposed by law, or (2) a phased implementation in which different dates for compliance are set for qualifying entities, depending on the number of their employees and whether the employer is a state entity, or in which different dates are set for degrees of compliance.

Qualifying employers

The selection of qualifying employers determines the scope of the implementation of gender pay enforcement mechanisms, i.e. it is the nature of the selection that determines how much of the workforce is included in the mechanism's ambit. Both the size of the employer in terms of number of employees in its workforce and whether the employer is public or private were used as qualifying criteria in the countries included in this research. The workforce size requirement is as few as 10 employees in Sweden, and as many as 500 employees in Germany. However, if the countries are looked at in terms of workforce size, it appears that those with smaller workforces (e.g. Sweden, Iceland, Denmark and Finland) have lower thresholds for implementation of the mechanisms that apply. There are exceptions, such as France, which has a workforce comparable to that of the UK and Germany, but has a low threshold of 50 employees. However, as a general rule, countries with larger workforces have higher thresholds, e.g. the UK (250 or more), Australia (100 or more) and Germany (500 or more). The rationale behind a higher threshold is to prevent smaller employers from being burdened with expensive and time-consuming reporting that may not have much influence on the gender pay gap[17], while lower thresholds are aimed at covering a greater proportion of the workforce. Some countries with lower thresholds have instituted a phased implementation, from larger to smaller employers, presumably to allow smaller employers more time to comply with the requirements. Most countries do not differentiate between public and private employers. However, Austria and Australia limit compliance to the private sector. One reason for this differentiation could be pre-existing reporting mechanisms, such as detailed annual reports, and collective bargaining. Another factor to consider is that legislation may specifically protect private sector pay information under data-protection and privacy laws.[25] In such cases, employers are prohibited from releasing pay information, and employees have no legal duty to do so.[25] The third part of this requirement is the existence of other qualifying criteria besides the above two, e.g. an obligation to publish management reports (Germany), being subject to federal regulation of employment matters or a federal contractor (Canada), or companies not being required to prepare annual reports (Norway). The overarching rationale for additional criteria would be to include certain types of companies in the mechanisms, e.g., in Canada, the federal government ensures that all companies with which it contracts are compliant with the mechanisms, while, in Germany, only certain types of companies have to produce management reports that enable deeper analysis of the operations of the company.

Reporting as a transparency mechanism

The use of a gender pay report was compared in the form of reporting with which qualifying employers must comply (report, survey, audit, etc.), the reporting period, and to whom the report must be provided. In France, employers only have to publish information on the size of the gender pay gap on their website. In Denmark, reporting takes the form of information submitted to Statistics Denmark; the data are then gender-segregated and returned to the employer, and employees also have access to this information. Finland uses an equality plan that is made available to employees, while Iceland requires certification of compliance with an Icelandic Standard with regard to the gender wage gap. By comparison, in India, companies are mandated to compile a register of employee data, including, inter alia, the number of men and women employed, their remuneration, and a breakdown of the components of remuneration, but there is no obligation to report this information. Similarly, Peru has, since 2017, required that companies compile detailed tables of employee categories and functions, including salary information. Finally, in Japan, companies must perform research on the gender pay gap and submit an action plan based on that research, which details how they will address the pay gap.

In terms of the frequency of using the reporting mechanisms, 2 of the 16 countries under study require reporting every 3 years, 7 require annual reporting, and 4 require biannual reporting. Of the remaining three countries, two do not have reporting requirements, and the last country only requires that a once-off plan be reported. While increased frequency has the disadvantage of increased administration to collect the required data, it is beneficial in ensuring that equitable gender pay practices are upheld.

Another important consideration in the present study was the target of the reporting mechanism, that is, whether it is internal or external to the

organisation. Internal transparency relates to information available to an organisation's employees within the organisational system. External transparency may serve various audiences, namely trade unions, regulatory bodies such as government, and the public. It was found that six countries require some form of external reporting, such as on the organisation's or public institution's website, or in the federal gazette. In four countries, the information is only available to local union representatives or the employer's works council. In Finland, the information need only be communicated to employees. In Austria, if there is no works council, the information must be made available directly to employees. However, in Peru and India, the information need not be communicated at all, but must be available for inspection by the appropriate authorities.

Indicators or measures used in different countries

What is actually reported on is crucial to understanding the nature of the gender wage gap in a particular country. The gap may be present in only certain occupational levels or industries, or within different components of the remuneration package. Identifying the source of the gap in a given employer's remuneration is key to formulating a response to reduce it. One common measure reported on is the gender pay equity objectives and policies in place at the employer. This measure is usually coupled with an analysis of the effects or results of such policies (Australia, Germany, Norway). Other countries have to report on the actual remuneration and bonuses paid to male and female employees, as performance bonuses and variable pay may contribute to the gender wage gap to a greater extent than base pay.[15] Another indicator is the number of men and women in each occupational level, which illustrates the distribution of men to women among the various jobs at an employer, or among different industries in a country, e.g. more women in health care and more men in mining. Australia has the additional requirement that each employer report on whether a remuneration gap analysis has been performed. This requirement encourages the employer to take a proactive stance in identifying and remedying internal gender wage gaps using internal statistics and data within the entity and to try to resolve any gaps identified before outside forces compel them to do so.

Penalties for non-compliance

Penalties for non-compliance should be sufficiently significant to be a deterrent. Non-compliance can refer to not reporting or to non-compliance with the obligation to pay women the same as men for work of equal value or the same work. The countries under study were found to have different approaches in this regard, ranging from no sanctions to public naming and shaming and financial penalties. In Australia and the UK, legislation does not provide for any sanction. In Australia, non-compliant employers could be named and shamed by the Workplace Gender Equality Agency, and the employer could be precluded from government tenders. In Sweden, an employee who is paid less than a comparable employee can raise a claim for equal payment. In Germany, failure to file the required gender equality report could be considered tantamount to not having filed an annual report, which could, but probably will not, lead to a fine. Iceland, Belgium, Denmark and Canada provide for administrative fines in their legislation. Canada and Iceland fine organisations fixed amounts for not filing reports or failure to obtain certification, while Belgian legislation provides for fines of variable amounts, depending on the number of employees affected by the non-compliance.

Mechanisms currently used in South Africa

The Constitution

The foundation of South African law is the Constitution of the Republic of South Africa of 1996 (the Constitution).[26(s1(c))] In Chapter 2, the Bill of Rights states that neither the state, nor any person, may unfairly discriminate directly or indirectly against anyone on the grounds of, inter alia, gender or sex.[26(s9(3),(4))] Furthermore, the state is obliged to enact national legislation 'to prevent or prohibit unfair discrimination'[26(s9(4))]. This obligation must be examined in light of South Africa's ratification of the ILO's Equal Remuneration Convention of 1951.[27]

National legislation

In terms of the *Promotion of Equality and Prevention of Unfair Discrimination Act* (PEPUDA)[28], which came into force on 15 January 2003, a complainant may allege a case of unfair discrimination, which, if based on any of the prohibited grounds listed in this Act, including gender, would necessitate the respondent proving that the discrimination was fair[28(s13)]. If the complainant succeeds in her or his claim, the presiding officer of the Equality Court 'may make an appropriate order in the circumstances, including an award of damages in favour of the complainant or an order directing that specific steps be taken to stop the unfair discrimination'[28(s21(2))]. This is an after-the-fact remedy, rather than a preventative measure, and is not limited to the workplace. However, the PEPUDA will only apply in instances in which the *Employment Equity Act*[29] (EEA), discussed below, does not apply[30]. Thus, both Acts have circumscribed areas of application.

Section 28 of the PEPUDA identifies the duty and responsibility of all persons to eliminate gender discrimination and to promote equality.[28(s28(3)(a))] This responsibility must be carried out via audits, promulgating appropriate laws, developing progressive policies, implementing codes of good practice, and instituting viable action plans.[28(s28(3)(b))] Finally, the Act contains a Schedule listing several practices occurring in different sectors 'which are or may be unfair, that are widespread and that need to be addressed'[28(s29(1))], ensuring that, in the appropriate circumstances, legislative and other measures are taken[28(s29(2))]. One of the practices identified in the 'Labour and Employment' section is '[f]ailing to respect the principle of equal pay for equal work'[28(Sch1(1)(c))]. Thus, the practice of not paying women equally to men for work of equal value has been identified as an unfair practice that the state has a duty to address.

In support of addressing the EEA[29], which came into force on 1 December 1999, was amended on 1 August 2014. Specifically, Section 6(4) states that '[a] difference in terms and conditions of employment between employees of the same employer performing the same or substantially the same work or work of equal value that is directly or indirectly based on [gender], is unfair discrimination'[29(s6(4))] – the position already taken by the Labour Court prior to these amendments[31].

The Regulations to the EEA[32] provide clarification on the definition of work of equal value, how to apply Section 6(4) of the EEA, and how to assess whether work is of equal value. Further to promoting gender wage equality, designated employers have a duty to submit an annual report (Form EEA2) to the Director General of the Department of Labour, including information regarding the workforce profile, numerical targets, and goals and skills development, but this information does not include employee pay information.[29(s21(1))]

The EEA defines a designated employer as: (1) a person employing at least 50 employees; (2) a person employing fewer than 50 employees, where such person has a total annual turnover equal to or exceeding that of a small business in terms of Schedule 4; (3) a municipality; (4) an organ of state, but not the National Defence Force, National Intelligence Agency, or the South African Secret Service; and (5) an employer that has been appointed a designated employer in terms of the EEA by virtue of a collective agreement in terms of Section 23 or 31 of the *Labour Relations Act 66 of 1995.*[29(s1)] A designated employer must also submit an Income Differential Statement (IDS – EEA4) that details the average remuneration and benefits received by employees on each occupational level of the employer's workforce – segregated by gender, race, and nationality, which must be submitted to the Employment Conditions Commission (ECC)[29(s27(1)), 32(reg12(1))]. Additionally, employers must provide remuneration details divided into fixed/guaranteed and variable portions of the highest paid employee in each subcategory of each occupational level, except at the lowest occupational level, where the remuneration amounts of the employee with the lowest total remuneration must be provided (Section D).[32(EEA4)] Thus, the actual data reported in this statement are a comparison of the total remuneration per occupational level in subcategories, and, as such the statement does not enable comparison of total remuneration of men to that of women for a given level. It only reports on subcategories, such as total remuneration of white women compared to that of African men. It is possible to derive those data from the statement, but it shows the more intersectional nature of the data that the Department of Labour is seeking.

The section titled 'Reasons for Income Differentials' requires employers to identify the reasons for income differentials at each occupational level. However, the employer is not required to specify to which income differential(s) related to gender, race or nationality the reason(s) apply.

It must be noted that the EEA4 is not a public document, in contrast to the EEA2 form, which may be requested by the public from the Department of Labour.[32(reg10(11))] However, during collective bargaining, the parties thereto may request the information contained in an IDS for the purpose of collective bargaining.

The purpose of an IDS is reflected in Section 27(2) of the EEA, being to determine whether there are 'disproportionate' income differentials, or whether there exists unfair discrimination in terms of Section 6(4).[29(s27(2))] The EEA protects individual employees' privacy by prohibiting the ECC from disclosing any identifying personal information.[29(s27(5))] If an employer is found to have unfairly discriminated against an employee, the Labour Court has the power to award compensation to the employee, to direct the employer to take steps to prevent the same unfair discrimination or similar practice occurring in future, or to direct an 'undesignated' employer to comply with Chapter III of the EEA, including an order to provide an IDS.[29(s50(2))] The court will probably order that the affected employee's remuneration be increased to that of the comparator.[33] Additionally, failure to submit the IDS could result in a fine being imposed on the employer.[29(ss21(4B),27(1))] If the employer has had no previous contravention, the maximum fine that may be imposed is the greater of ZAR1 500 000 or 2% of the employer's turnover.[29(sch1)] The fine increases with the number of previous contraventions the employer has committed.[29(sch1)]

Governance codes

According to the King IV Report on Good Governance (2016), a company's board must approve reports on and the implementation of its remuneration policy, which should reflect that 'the organisation remunerates fairly, responsibly and transparently'[34]. The King IV Report is binding on companies listed on the JSE Limited.[35(s3.84(c))] The policy and report must be tabled at the company's annual general meeting for a separate, non-binding advisory vote by shareholders.[35(s3.84(k))] If 25% or more of shareholders vote against either document, the company must invite dissenting shareholders to provide reasons for their vote.[35(s3.84(k))] Additionally, the annual report must indicate the remuneration of the company's directors.[35(s8.62-3),36(ss30(4)-(6))] Therefore, gender is not specifically targeted in the JSE's requirement for reporting of directors' remuneration, which information could be used to determine the existence of a gender pay gap at director level.

Recommendations

Despite ILO Conventions No. 100 (Equal Remuneration) and No. 111 (Discrimination: Employment and Occupation) having been in existence for over 60 years, the gender pay gap is still a concern in industrialised countries, and an even greater one in developing countries.[5] Taking into consideration that 'policies targeting the gender wage gap must be adjusted to account for the different experiences of women in different parts of the wage distribution'[1] and the phased approach highlighted in the global comparison, we make the following recommendations.

The first requirement is the passing of laws at regional and national level that place a duty on employers to give men and women equal remuneration for the same or similar work, or work of equal value.[25] South Africa achieved this requirement with the implementation of Section 6(4) of the EEA. The criteria to determine designated employers, as stipulated in the EEA, follow global trends, reflecting appropriateness for the size of South Africa's workforce. Inclusion of both public and private organisations in mandatory reporting aids identification of patterns in gender wage gaps, and could be used to formulate targeted policies to close the gap. However, the EEA addresses mainly pay discrimination at the individual job level, and addressing the issue through litigation is costly and intricate for affected individuals.

Whilst the IDS serves as a preliminary mechanism to flag inconsistencies regarding a number of intersectional wage differentials at aggregated level and the gender wage gap receives specific attention, the format of the data only enables national comparison in pre-determined occupational categories. Legislation specifying duties of employers and penalties for non-compliance, as seen in Belgium and Sweden[37], is the preferred method to promote pay transparency and equality[25], and it is in this area that South Africa can improve. Specifically, the EC Recommendation[18] provides a benchmark of four mechanisms that South Africa could employ.

The first mechanism is the strengthening of pay reports by further refining an annual reporting obligation on gender-segregated average remuneration for medium and large companies already targeted by the EEA. Supplementary tables 1–3 provide a description of various measures or indicators employed globally. A selection of indicators of the UK, Australia and Canada would provide the widest coverage of gender-relevant aggregated data. Reports should be lodged with the ECC as a first step, with consideration to phasing in the publishing of company data more publicly in future, once implementation problems have been resolved. Reporting that is too generic, as is presently the case with the IDS in South Africa, conceals structural inequalities, leaving policymakers to apply a one-size-fits-all approach instead of targeted solutions. Data could remain anonymous through the inclusion of a criterion of a minimum number of employees per reporting category, e.g. there must be at least five employees at a specific level in order for that level to be reported on[17]. Whilst the complexities in reporting the annualised value of non-vested share incentive schemes are acknowledged (these are currently included in the IDS only as part of the variable remuneration figure), inclusion thereof in pay reports as a separate category of reporting is essential in identifying gender-based differences in performance-related pay, and should be phased in in the near future. A rapid gain could be made if the JSE expanded its interpretation of King IV's stipulation that 'the organisation remunerates fairly, responsibly and transparently' by mandating gender gap reporting in annual reports, through the Listing Requirements, for all employees in the bottom, middle and top of the wage distribution of an organisation, including directors. Such gender-specific reporting covering the main infliction points in the wage distribution could inform targeted policies to close the gap in listed companies.

The EU recommendation that employers regularly communicate pay report information to employees or trade unions (or other representatives) is absent in South African legislation, and should be considered. The pay report mechanism will further be strengthened if the parties analyse such a pay report and raise queries regarding gender pay discrepancies. In this regard, employees should be educated in remuneration principles and practices in order to limit misinformed queries and ungrounded discontent and disputes. A pay report that specifically provides gender-segregated pay information could be used as a precursor to an employee's right to request pay information.

The second mechanism, an employee's right to query another employee's pay, could be problematic in South Africa, due to a person's right to privacy and confidentiality[26(s14)], as also identified in the EC report. The EEA prohibits the ECC from disclosing any information pertaining to individual employees or employers, which would include details of the IDS.[29(s27(5))] This restriction is especially pertinent where companies have very few employees in the respective occupational level or gender grouping.

Another consideration is that the *Protection of Personal Information Act*[38] (POPI) regulates the processing and distribution of personal information. However, the relevant sections of the POPI Act concerning employee information are not yet in force, and are thus not further discussed here.[39] With reference to the *Basic Conditions of Employment Act* (BCEA)[40], an employee may disclose his or her own remuneration[40(s78(1)(b)),41]. An employer may not contractually prohibit an employee from disclosing this information.[40(s79(2)),41] Thus, an employee may obtain a comparator's salary information directly from the comparator.[33(p.829),40(s78(1)(b))] However, an employee may not disclose information regarding another employee's salary if contractually bound to keep such information confidential.[42] In this respect, the use of the *Promotion of Access to Information Act*[43] (PAIA) may be of help in securing this information. Section 63(2)(f) thereof states that a record may not be refused if the information – including remuneration – relates to the official of a private body (see Section 34(2)(f) for a similar provision regarding public bodies).[43] An 'official' is defined as 'any person in the employ (permanent or temporary, full time or part time) of the public or private body'[43(s1)]. Thus, an employee would, in terms of PAIA, be able to obtain the remuneration information of another employee upon such request.[44(8-106)] However, as the duty to disclose is subject to other legislation, such as

the EEA[43(s5)], salary information would more likely have to be requested from the employer and not from the ECC, which could be prohibited from disclosing such information, even in terms of PAIA.

The EC has also found several other barriers to implementing this employee right, such as cultural sensitivity against disclosure, the administrative and cost burden thereof, and contractual non-disclosure clauses.[17] It must also be considered that an employee needs to have the will and ability, specifically financially, to enforce this right, but having the support of a trade union could alleviate the burden of such a potentially cumbersome and costly process.[25] Despite these issues, the legislature should seek to give an employee this right, especially when the employee is attempting to prove a claim in terms of Section 6(4) of the EEA, in which case the correct comparator is needed.

The third proposed measure is a pay audit at the level of the employer.[18] The key differentiator between an audit and a pay report is the analysis of pay gaps found in the former.[17] Such analysis[45] could enable the identification of the problem and the development of targeted countermeasures[17]. The implementation of a pay audit could provide useful data, but has the downside of costs in terms of time and expert fees to analyse the data and devise measures to close gaps.[17] However, in order to meet the requirements of fair, responsible and transparent remuneration practices, as stipulated in King IV, companies with a strong governance ethos would benefit from mandatory pay audits. Smaller companies, however, may not have the means with which to conduct such audits. Thus, legislated comprehensive pay audits may not be ideally suited to the current South African context; however, in following the principle of a phased approach, we recommend that gender pay audits be legally mandated for large and listed companies as an initial step.

The final recommended measure is ensuring that companies discuss, as a separate issue, equal gender pay, including pay audits, during collective bargaining.[17,18] The effectiveness of such a measure depends on the development of collective bargaining in a specific sector.[17] The more developed the collective bargaining structures are and the more unionised the workforce is, the easier it will be to implement this measure.[17] In South Africa, collective bargaining almost always occurs in the government, resources and industrial sectors, while small companies, in general, may have few or no worker representatives.[46] The inclusion of gender pay equality as a collective bargaining issue may be legally mandated or a soft law duty on the parties to collective bargaining.[17] This measure will also be more effective if the state was allowed to intervene during the collective bargaining process.[17] This measure carries the advantage of creating an awareness of gender pay issues at a higher sectoral level, but could impose onerous duties on the parties to collective bargaining, which may already be burdened with setting remuneration for different levels of workers in different companies.[17] Furthermore, trade unions may view this measure as a threat to their autonomy[17], especially when considering the *Labour Relations Act*[47], wherein no role is identified for the state in collective bargaining processes. Due to the conflicts that have arisen around collective bargaining in South Africa, we suggest that a soft law stipulating that companies discuss equal pay, including pay audits, during collective bargaining be introduced into the King codes as a matter of good remuneration governance.

Lastly, an important component of effective transparent mechanisms is penalties for non-compliance with stipulations. This component necessitates diligent monitoring for non-compliance and enforcement of penalties. It is recommended that a financial penalty be levied for unjustifiable and stagnant gender pay gaps among the employees of the same employer – one that is sufficient to act as a deterrent to non-compliance (as a fixed amount per period of non-compliance), as is done in Iceland, or be calculated per employee found to have been discriminated against, as is done in Belgium. Penalties should thus promote compliance with gender pay legislation and transparency mechanisms, and ultimately disincentivise discriminatory pay practices.

Conclusion

The World Economic Forum estimates that it may take another 202 years before the gender economic gap is closed.[48] Embedded in this gap is the gender wage gap. South Africa has made great strides towards addressing gender inequality, but purposeful momentum towards this goal has waned. The proposals in this paper are aimed at reigniting debate and motivating action to attain gender wage equality by amending policies towards transparency.

Acknowledgements

We thank Prof. Nicola Smit, Prof. Hugo Pienaar and Ms Soula Proxenos for their helpful comments.

Authors' contributions

A.B. conceptualised the study and co-wrote the article. S.B. collected data and co-wrote the article.

References

1. Mosomi J. Distributional changes in the gender wage gap in the post-apartheid South African labour market. WIDER Working Paper 2019/17 [document on the Internet]. c2019 [cited 2019 Jun 16]. Available from: https://www.wider.unu.edu/publication/distributional-changes-gender-wage-gap-post-apartheid-south-african-labour-market
2. International Labour Organization (ILO). Global wage report 2018/19: What lies behind gender pay gaps. Geneva: ILO; 2018. Available from: https://www.ilo.org/wcmsp5/groups/public/---dgreports/---dcomm/---publ/documents/publication/wcms_650553.pdf
3. Rubery J, Grimshaw D, Figueiredo H. How to close the gender pay gap in Europe: Towards the gender mainstreaming of pay policy. Ind Relat J. 2005;36(3):184–213. https://doi.org/10.1111/j.1468-2338.2005.00353.x
4. African Development Bank Group. Africa gender equality index [webpage on the Internet]. c2015 [cited 2019 Jul 16]. Available from: https://www.afdb.org/en/topics-and-sectors/topics/quality-assurance-results/gender-equality-index/
5. World Economic Forum. The global gender gap report 2018 [document on the Internet]. c2018 [cited 2019 May 10]. Available from: http://www3.weforum.org/docs/WEF_GGGR_2018.pdf
6. Statistics South Africa (Stats SA). General household survey 2018. Statistical release P0318. Pretoria: Stats SA; 2019. Available from: http://www.statssa.gov.za/publications/P0318/P03182018.pdf
7. Statistics South Africa (Stats SA). Living conditions of households in South Africa: An analysis of household expenditure and income data using the LCS 2014/2015. Statistical release P0310. Pretoria: Stats SA; 2017. Available from: http://www.statssa.gov.za/publications/P0310/P03102014.pdf
8. South African Institute of Race Relations. South Africa's family fabric. FreeFACTS. 2018;5(5):1–5. Available from: https://irr.org.za/reports/freefacts/files/free-ff-2014-september-2018.pdf
9. 10X Investments. 10X South African retirement reality report 2018 [document on the Internet]. c2018 [cited 2019 Jun 16]. Available from: https://cdn2.hubspot.net/hubfs/3390004/10X%20Document/10X%20SA%20Retirement%20Reality%20Report%202018.pdf
10. Soudien C, Reddy V, Woolard I. Poverty & inequality: Diagnosis, prognosis and responses. Cape Town: HSRC Press; 2019.
11. Vetten L. Reconstruct. The Sunday Independent. 2000 March 12. Available from: https://www.csvr.org.za/media-articles/all-csvr-in-the-media/2173-the-economics-of-domestic-violence
12. Gouws A. Fairness in the family. In: Bosch A, editor. South African Board for People Practices women's report 2017. Johannesburg: South African Board for People Practices; 2017. p. 8–12. Available from: https://sabpp.co.za/wp-content/uploads/2019/05/SABPP_Womens-Report-2017_Fairness-in-relation-to-women-at-work.pdf
13. Peetz D. Regulation distance, labour segmentation and gender gaps. Cambridge J Econ. 2015;39:345–362. https://doi.org/10.1093/cje/beu054
14. Greiner B, Ockenfels A, Werner P. Wage transparency and performance: A real-effort experiment. Econ Lett. 2011;111(3):236–238. https://doi.org/10.1016/j.econlet.2011.02.015
15. Rubery J, Grimshaw D. The 40-year pursuit of equal pay: A case of constantly moving goalposts. Cambridge J Econ. 2015;39:319–343. https://doi:10.1093/cje/beu053

16. DLA Piper. Gender pay reporting: An international survey [document on the Internet]. c2018 [cited 2019 Jun 16]. Available from: https://www.dlapiper.com/~/media/files/insights/events/2018/11/international_gender_pay_gap_report_2018_us.pdf

17. Veldman A. Pay transparency in the EU: A legal analysis of the situation in the EU member states, Iceland, Liechtenstein and Norway. Brussels: European Commission; 2017. https://doi.org/10.2838/148250

18. European Commission. Commission recommendation of 7.3.2014 on strengthening the principle of equal pay between men and women through transparency [document on the Internet]. c2014 [cited 2019 May 11]. Available from: http://ec.europa.eu/smart-regulation/impact/ia_carried_out/docs/ia_2014/c_2014_1405_en.pdf

19. Rasmussen K. General approaches to data quality and Internet-generated data. In: Fielding N, Lee RM, Blank G, editors. The SAGE handbook of online research methods. 2nd ed. London: SAGE; 2008. p. 79–97. https://doi.org/10.4135/9780857020055.n5

20. United Nations. Member states [webpage on the Internet]. No date [cited 2019 Jul 16]. Available from: https://www.un.org/en/member-states/#gotoP

21. Allen GN, Burk DL, Davis G. Academic data collection in electronic environments: Defining acceptable use of internet resources. MIS Quart. 2006;30(3):599–610. https://doi.org/10.2307/25148741

22. Carvajal TG, Hernandez-Salgado L. Spotlight on the gender pay gap in Latin America [webpage on the Internet]. c2018 [cited 2019 Jun 15]. Available from: https://www.lexology.com/library/detail.aspx?g=ab0f2b11-6688-416f-807a-dfe3daa851b4

23. Lichuma W. No rights without remedies: Gender equality and secularism in Africa. Conscience. 2017;38(4):20–21.

24. Hall J. Defending human rights in Africa. Africa Conflict Monitor. 2015;2015(10):11–17.

25. Rubery J, Koukiadaki A. Closing the gender pay gap: A review of the issues, policy mechanisms and international evidence [document on the Internet]. c2016 [cited 2019 May 10]. Available from: https://www.ilo.org/wcmsp5/groups/public/---dgreports/---gender/documents/publication/wcms_540889.pdf

26. Constitution of the Republic of South Africa, 1996.

27. International Labour Organization. Ratifications for South Africa [webpage on the Internet]. No date [cited 2019 May 06] Available from: https://www.ilo.org/dyn/normlex/en/f?p=NORMLEXPUB:11200:0::NO::P11200_COUNTRY_ID:102888

28. Promotion of Equality and Prevention of Unfair Discrimination Act 4 of 2000, South Africa.

29. Employment Equity Act 55 of 1998, South Africa.

30. Cooper C, Lagrange R. The application of the Promotion of Equality and Prevention of Unfair Discrimination Act and the Employment Equity Act. ILJ. 2001;22:1532–1545.

31. Pienaar H. The meaning of equal value in terms of pay. In: Bosch A, editor. South African Board for People Practices women's report 2015. Johannesburg: South African Board for People Practices; 2015. p. 18–22. Available from: /https://sabpp.co.za/wp-content/uploads/2019/05/SABPP_Womens-Report-2015_Equal-pay-for-equal-value.pdf

32. Employment Equity Regulations, 2014, South Africa.

33. Laubscher T. Equal pay for work of equal value – a South African perspective. ILJ. 2016;37:804–830.

34. Institute of Directors Southern Africa. King IV report on corporate governance for South Africa 2016 [document on the Internet]. c2016 [cited 2019 Jun 16]. Available from: https://c.ymcdn.com/sites/iodsa.site-ym.com/resource/collection/684B68A7-B768-465C-8214-E3A007F15A5A/IoDSA_King_IV_Report_-_WebVersion.pdf

35. JSE Limited. JSE Listing requirements: Service issue 26 [document on the Internet]. c2018 [cited 2019 Jul 16]. Available from: https://www.jse.co.za/content/JSERulesPoliciesandRegulationItems/JSE%20Listings%20Requirements.pdf

36. Companies Act 71 of 2008, South Africa.

37. United Nations Economic Commission for Europe. The gender pay gap in UNECE countries: What has changed since 2000? [webpage on the Internet]. c2018 [cited 2019 Jul 06]. Available from: https://www.unece.org/info/media/news/statistics/2018/the-gender-pay-gap-in-unece-countries-what-has-changed-since-2000/doc.html

38. Protection of Personal Information Act 4 of 2013, South Africa.

39. Michalsons. POPI commencement date or POPI effective date starts the clock [webpage on the Internet]. c2019 [cited 2019 Jun 17]. Available from: https://www.michalsons.com/blog/popi-commencement-date-popi-effective-date/13109

40. Basic Conditions of Employment Act 75 of 1997, South Africa.

41. Matsi and JP Hugo Residence CC t/a Hoffe Park Accommodation Centre (2013) 34 ILJ 1018 (CCMA).

42. Salt L. Employees cannot contract out of right to disclose salary information [webpage on the Internet]. No date [cited 2019 May 14] Available from: https://www.labourguide.co.za/most-recent/1845-employees-cannot-contract-out-of-right-to-disclose-salary-information

43. Promotion of Access to Information Act 2 of 2000, South Africa.

44. Pretorius JL, Klinck ME, Ngwena CG. Employment equity law: Service issue 18. Durban: LexisNexis; 2018.

45. Boninelli I. The gender pay audit – practical steps for multi-national and local companies. In: Bosch A, editor. South African Board for People Practices women's report 2015. Johannesburg: South African Board for People Practices; 2015. p. 7–12. Available from https://sabpp.co.za/wp-content/uploads/2019/05/SABPP_Womens-Report-2015_Equal-pay-for-equal-value.pdf

46. Wagner I. Certified equality: The Icelandic Equal Pay Standard [webpage on the Internet]. c2018 [cited 2019 Jun 16]. Available from: https://www.researchgate.net/publication/329371051_Certified_Equality_The_Icelandic_Equal_Pay_Standard

47. Labour Relations Act 66 of 1995, South Africa.

48. World Economic Forum. Closing the gender gap [webpage on the Internet]. Geneva: World Economic Forum; 2019 [cited 2019 Jul 16]. Available from: https://www.weforum.org/projects/closing-the-gender-gap-gender-parity-task-forces

AUTHORS:
Sikanyiso Masuku[1]
Sharmla Rama[2]

AFFILIATIONS:
[1]Institute for Democracy, Citizenship and Public Policy in Africa, University of Cape Town, Cape Town, South Africa
[2]Society and Social Change Cluster, School of Social Sciences, University of KwaZulu-Natal, Pietermaritzburg, South Africa

CORRESPONDENCE TO:
Sikanyiso Masuku

EMAIL:
sikhanyiso.masuku@gmail.com

DATES:
Received: 08 Apr. 2019
Revised: 04 Sep. 2019
Accepted: 17 Sep. 2019
Published: 26 Mar. 2020

HOW TO CITE:
Masuku S, Rama S. A case study of government and civil societies' collaboration and challenges in securing the rights of Congolese refugees living in Pietermaritzburg, KwaZulu-Natal, South Africa. S Afr J Sci. 2020;116(1/2), Art. #6210, 6 pages. https://doi.org/10.17159/sajs.2020/6210

ARTICLE INCLUDES:
☒ Peer review
☐ Supplementary material

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

EDITOR:
Hester du Plessis

KEYWORDS:
migration, state-assisted integration, human rights, refugee policy implementation, bureaucratic efficiency

FUNDING:
None

A case study of government and civil societies' collaboration and challenges in securing the rights of Congolese refugees living in Pietermaritzburg, KwaZulu-Natal, South Africa

Several South African focused studies have identified the numerous challenges faced by refugees in securing their legal rights to employment, education, health care, etc. There is a need therefore to investigate the extent to which such challenges are conterminous with bureaucratic institutions (such as government departments) limitations in fully implementing the refugee policy. In cognisance of the many shortcomings associated with state-assisted integration models, we examined how bureaucratic efficiency (in the provisioning of refugee rights) can be achieved, i.e. the feasibility of instituting dependency partnerships between the state and civil society. A case study based interpretive research design technique was adopted, drawing from one focus group discussion with Congolese refugees and face-to-face in-depth interviews with three purposively selected NGO representatives. We used two theoretical perspectives, namely the theory of monopolisation and Weber's theory on bureaucracy, to analyse how bureaucratic arrangements can negatively impact on the implementation of the refugee policy and consequently on the refugees' quality of life and standard of living in their host country. We identified that primary cultural factors, amongst others, unruly practices, social closure, and institutional biases widen the chasm between the formulation of a progressive refugee policy and its efficient implementation. The setting up of human rights education interventions and dependency partnerships is recommended as a means of improving bureaucratic efficiency in the transfer or implementation of refugee social protections.

Significance:

- The originality of this paper emanates from its conceptualisation of the state as a bureaucratic institution whose efficiency in implementing policies can be undermined by the existence of not only structural but also primary cultural factors. Refugees' challenges are thus conceptualised as emanating from a failure to achieve bureaucratic efficiency.
- Multisectoral approaches (dependency partnerships) are suggested as an alternative to a purely vertical top-down model whereby a single bureaucratic institution (often prone to institutional biases and other challenges) is primarily responsible for implementing social protection policy.
- Apart from not providing a hypothesis-based analysis of refugee deprivations, other studies on refugees in South Africa and the region do not examine such deprivations from an institutional standpoint.

Introduction and background

Chambers and Kopstein[1] acknowledge the importance of public and private sector partnerships in policy implementation as well as in averting refugee deprivations. In this paper, we examine how non-governmental organisations (NGOs) and stakeholders can augment governments' efforts in the refugee social protection discourse[2,3], i.e. what Berten and Leisering[4] term 'inter-organisational exchanges'. We examine refugees' rapport with civil society through analyses of the findings from interviews with a sample of civil society groups as well as Congolese refugees in South Africa.

The existential limitations in the adoption of a collective approach to the transfer of refugee social protections[5] (multisectoral initiatives) are investigated. This inquiry is rationalised by the existential challenges to refugee groups' access to a spectrum of rights and services in South Africa.[6] These challenges are hypothesised as principally originating from the deprivations associated with vertical top-down interventions or state-assisted integration models, i.e. institutional biases, normative forms of exclusion and so forth.[5,7]

As found in Germany, through their national law on integration, as well as in Sweden, Italy, the Netherlands and Norway, through state-assisted integration programmes, multisectoral social initiatives often reduce refugees' deprivations.[8,9] Bottom-up approaches to refugee social protection have been discernible through civil society groups' principal role in organising human rights education in places such as Zambia, Sudan, Uganda and Kenya.[5,10] We therefore examined how NGO's greater access to grassroots communities means they are better placed to facilitate what Berten and Leisering[4] refer to as the 'bottom-up theorisation' of social protection policy. This paper is dissimilar to other studies on refugees in South Africa, in that we investigate the nexus between the existential challenges in forging inter-organisational exchanges (multisectoral partnerships) and the arising limitations in the transfer as well as implementation of refugee policy.

State-assisted integration

Governments are obliged to play a pivotal role in securing refugees' social security rights. Government policies targeting inequality and favouring social solidarity can promote integrated societies. A considerable impediment to

state-assisted integration in South Africa has been the issue of exclusionary proclamations by the government through the existence of predatory/unpredictable immigration systems.[11] This includes the prohibition on the right to work for asylum seekers (finally overturned in 2002), the nationwide closure of refugee reception offices as well as the denial of the right to family for forced migrants in the country. As another challenge to state-assisted integration, there has also been little coordination in the implementation of refugee policy in South Africa.[12] These declarations have infringed not only on refugees' attempts to integrate into their host countries but also on their rights to the preservation of their dignity as enshrined in Article 12 of the 1951 Convention and Article 16 of the Universal Declaration of Human Rights.[13,14]

This situation has been worsened by some citizens' systematic blockade of refugee groups' liberties. Although we identify refugee groups service exclusion from government departments as a derivative form of exclusion,[15] such social closure is also conceptualised in the paper as a multidimensional phenomenon that state-assisted integration cannot remedy on its own.

Refugee policy enactment: Locating the civil society

Civil society organisations across the globe are often proactive in facilitating refugees' access to both bridging and bonding forms of social capital.[16] Civil society organisations are thus particularly important in minimising refugees' deprivations through their role in the first steps towards developing a sense of self-identity (which does not necessarily emerge through engagement with the state). Civility, as defined by Harbeson et al.[17] advocates for equitable rights and entitlements within all human societies and it is through the collective efforts of civil society groups that this agenda has been pushed. Examples of where the model of civility has been successfully instituted include aid agency programmes designed to extend the rights to education, shelter and health care, i.e. through the Comprehensive Refugee Response Frameworks as implemented in Tanzania and Uganda.[18-20] Here we examine the model of civility in cognisance of how it promotes parity and equitable rights access for all.

Establishing the level of rapport between government and civil society is particularly important as it helps to ascertain the role of civil society as a stakeholder in human rights issues. Civil action groups formed by vulnerable groups (such as the Congolese refugees in this study) are a form of solidarism, symbolising a combined response to the challenges affecting them.[21] Fukuyama[22,23] argues that by coming together in civil associations, vulnerable groups become strong. As a means of mobilising power to enhance or defend a group's share of resources, Fukuyama[23] argues that, 'solidarism seeks to challenge the prevailing system of distribution'. Evidence of South African civil society organisations importance in cushioning refugees from social risks is provided by their pro-activeness during the perennial xenophobic attacks.[8]

Theoretical frameworks

Weber's[24] theory on bureaucracy identifies government as a rational legal authority responsible for the drafting and implementation of policy. It is from such a hypothesis that Weber developed his concept of the ideal bureaucratic organisation which is characteristically rational and efficient. Weber[24] argues that what Selznick[25] refers to as a 'depersonalisation of administrative relationships', although important in ensuring organisational efficiency and egalitarianism, is difficult to achieve. In his micro-theory of class and stratification, he argues that such a depersonalisation is not possible because legal rational authority is always at threat from traditional and historical prejudices. The balance between the three types of authority (traditional, charismatic and rational-legal) is what ultimately determines bureaucratic efficiency.[26] The argument here is that the rational legal authority in South Africa (*Refugee Act 130 of 1998*), regarded as the more forceful and effective form of authority, is constantly undermined by traditional authority, i.e. the counteractive actions of street level bureaucrats within state departments. Amongst other issues, this action compromises the successful transfer/implementation of refugee social protections in the country.

This raises concerns about how primary cultural forms of exclusion (debilitative sociocultural relations based on race, gender, citizenship or nationality) impact on the effectiveness of government (as a bureaucratic institution) in its mandate to fully implement refugee policy. Fraser[27], through his concept of unruly practices, argues that such practices in institutions often account for the 'gap between rules and their selective implementation'. Murphy[28] conceptualises the gap between rules and their prejudicial implementation as a derivative form of exclusion. Instead of operating autonomously then, 'dependency relationships' (minimal in both quantity and importance) should be sought between government and other organisations.[29] In examining how refugee policy implementation can be improved, this paper supports the view of the feasibility of inter-organisational exchanges.

In analysing the various ways through which refugees' access to a spectrum of services and commodity bundles is impeded, the study adopts the theory of monopolisation and exclusion.[28] This theory develops from the social closure tradition of Weber[30] and identifies conditions of exclusion as a multidimensional phenomenon, perpetuated on the grounds of an outside groups' deviance in race and other social attributes.[28] Through derivative forms of exclusion (exclusion based on racial, ethnic, religious, nationality or gender criteria), individuals can be disfranchised from accessing social protections and a host of other necessaries.[31] Deprivations and social closure can therefore be attributed to an absence of social, symbolic and cultural forms of capital amongst refugees.[32] We thus hypothesise the challenges in provisioning refugees' social protections as conterminous with existential biases, unruly practices, monopolisations, etc. within bureaucratic institutions.[7,32-34]

Methodology

We adopted an interpretive research design technique[35] and the case study is exploratory in terms of examining how a multisectoral approach could be adopted in containing refugees' deprivations and social exclusion[36]. In assessing the government's effectiveness as a bureaucratic institution, we utilised an approach whereby the refugee interview participants purposively consisted of the beneficiaries of the refugee policy under examination (section 24 permit holders). The need to interview government representatives was offset by the data obtained from the refugee participants on whether or not the state department services were indicative of an efficient bureaucratic institution. Their lived experiences and realities in South Africa give recognition to voices and experiences that are marginalised and devalued.

Through expert purposive sampling, we identified a Congolese representative association in Pietermaritzburg. We then utilised a homogeneous sampling technique to obtain eight focus group discussion participants from the association's membership. The use of only one nationality in the study was rationalised by how the Congolese refugees sufficed as an instrumental case in the examination of issues negatively impacting on the generality of refugees in South Africa.[37] The focus group discussion lasted a total of 90 minutes.[38] In selecting focus group discussion participants, the use of a non-probability homogeneous sampling method was necessitated because the study targeted candidates who shared similar traits or specific characteristics (Congolese refugees in South Africa).[39] Stratified purposeful sampling[40] was then used to sample four male and four female participants for the gender balanced focus group discussion. In having a gender-mixed focus group discussion, we identified power dynamics that came into play; in such instances, the facilitator moderated the proceedings by managing those who attempted to dominate the discussions. In selecting these participants, there was no focus on age although all participants were above the legal age of consent and had been in the country for a period not less than 10 years. For the purposes of an instrumental case study, a longer stay in South Africa ensured that the selected participants had a greater experience with the issues under discussion (this was then used to impress on the experiences of other refugees in the country).

In obtaining a sample from civil society, we utilised a purposive expert sampling technique. Participants were therefore experts purposively selected (based on their activism in refugee rights) and comprised one participant from the KwaZulu-Natal Christian Council and one from

Lawyers for Human Rights (LHR). In assessing the expertise of the organisations, we examined their scope, their programme's focus areas and overall social impact. We interviewed one participant from the civil society organisation representing Congolese refugees, the Congolese Refugees Association (CRA). Apart from these participants being active on provincial and national platforms, they were well informed about the issues affecting refugees.

Ethical clearance for the study was provided by the University of KwaZulu-Natal Humanities & Social Sciences Research Ethics Committee (reference number HSS/0810017D).

Results and discussion

Challenges to state-assisted integration

South Africa is signatory to a number of international conventions governing the well-being of refugee groups; establishing the role of government in inhibiting refugees' social disaffiliation particularly in accessing the available social protections is important. A beneficiary-based inquiry on the institutional efficiencies and deficits is therefore equally relevant and insightful in terms of understanding how social deprivations are exacerbated and experienced because of the actions or inactions of public servants and institutions. While refugees are defined in the Geneva 1951 Convention as groups or individuals who have left their own countries because of persecution or violence,[14] South Africa initially classifies such arrivals as asylum seekers, i.e. section 22 permit holders, with section 24 permits (refugee status) only being conferred after a lengthy (individualised) adjudication process. For that reason, refugees in South Africa (as guided by the *Refugee Act 130 of 1998*) were taken in this study to mean section 24 (refugee status) permit holders.

In South Africa, refugee legal rights are an Act of parliament enshrined in Regulation 27 (G), Regulation 27 (F), Regulation 27 (B) and Regulation 27 (D) of the *Refugee Act 130 of 1998* as well as Chapter 2, Articles 9 and 11 of the Bill of Rights. Weber's theory on bureaucracy and conceptualisation of government as a rational legal authority supports Congolese refugees' perceptions and beliefs that the host government is the principal custodian of their legal rights.[41] This study foregrounded how Congolese refugees anticipate the government to facilitate their integration and civil rights access in the country.

Although the legislative arms of government were progressive through liberal policy formulation, the study identified how such reforms were often undermined by the existence of an exclusionary traditional authority. Murphy[28] and Weber[24] submit that, despite the existence of progressive policy, group attributes (such as race, ethnicity, religion, language, nationality or citizenship) can be utilised to close off social and economic opportunities to perceived out-groups, i.e. traditional and historical prejudices can be a threat to the realisation of legal rational authority. South Africa's historical past of colonialism and apartheid has mutated into the perpetuation of racialised and discriminatory attitudes/actions post-apartheid. The ensuing ideological gaps between the *Aliens Control Act (No 96 of 1991)* and the *Refugee Act 130 of 1998* is a prime example of the disjunction between ideologies, action and authority.[12] The prejudicial attributes of traditional authority instigated through the unruly conduct of street-level bureaucrats (police, social workers, refugee status determination officers etc.), partly accounts for the gap between progressive policy formulation and its implementation.

Although Selznick[25] argues that there is a need for a 'depersonalisation of administrative relationships', the prevalence of corruption, and institutionalised prejudices and unruly actions is evidence of this failure[34,42]. Referred to by Brauns[26] as the intermission between legal rational authority and traditional authority, this has compromised the efficiency of government and its bureaucratic institutions to implement the refugee policy. The refugee participants indicated how this phenomenon was most prevalent amongst public servants or staff, and government's apparent failure to ensure the realisation of the *Batho Pele* [People First] principles.

Principal forms of exclusion and monopolisation can be enforced by the apparatus of the state through legal and coercive sanctions.[15,33,42] One civil society representative mentioned how the issues exacerbating refugees' deprivations and derivative exclusion were administrative. As opposed to integrating refugees within areas of commerce, certain councillors were closing off refugees based on their group attributes (race, ethnicity, language etc.).[32] One participant complained: 'There are some councillors who are saying we no longer want any foreign nationals to operate here.'

Through Murphy's[15] theory, such social exclusion and prejudices in defiance of pre-existing progressive laws, can be seen as being premised on the grounds of 'credentialism, racial, ethnic monopolisation and social exclusion'. Weber's micro-theory on class stratification also illuminates how exclusionary traditional and historical prejudices may subvert a progressive legal authority. Such a subversion may manifest through unruly practices, biases, monopolisations and exclusion. As a consequence of government's failure to effectively facilitate their integration, Congolese refugees indicated that they were deprived of equal access to health, education, shelter, etc. Such deprivations mainly emanated from structural or principal forms of exclusion.

Refugee participants also complained of victimisation within government departments:

> *Most refugees are being victimised by the SAPS, you will be at your place at night and the police officers will just pop in pretending that they are looking for illegal immigrants. So, when you open, they will start asking for papers for furniture, television licence etc. and if you don't have receipts, they will threaten to take you to the police station until you come and prove that you bought it in the shop. It can also happen at the place of work, they will come in the saloons and say all of you present your papers, whoever does not have they take him to a corner where you have to pay* [a] *bribe.*

The conduct described above in which workers within the public service disregard refugee policy edicts, reflects a situation whereby traditional authority incapacitates rational-legal authority and ultimately bureaucratic efficiency.[26] Khan et al.[43] conceptualises it as a form of unruly practice which impedes state-assisted integration models. The interviews therefore revealed how Congolese refugees expect the state to address the exclusionary nodes not only within their communities of residence but also within the civil service. Public service employees' ignorance of refugee policy was identified as an additional impediment to refugees' access to their legal rights. Other studies corroborate these findings and allude that the exclusionary actions of public service staff often exacerbate refugees' deprivations.[44,45] The prejudice and deprivations refugees face due to social exclusion were confirmed by the following responses:

> *When it comes to social services there is the belief that refugees are dirty, there are certain mind-sets that should be changed with regards to frontline service providers who deal with refugees.*

> *…in the trainings that we have, the government officials will come in the open and say we don't know these things, which means they are not alone they are a lot out there and maybe some refugees can be disadvantaged because somebody doesn't know.*

In addressing the above challenges, participants suggested a sensitisation programme for public servants due to their role as key players in the aversion of social exclusion both inside and outside public institutions. These sentiments are substantiated by other researchers who argue that issues of social inclusion require awareness campaigns and a change of consciousness.[46-47] Khan et al.[43] concur and argue that such a change of consciousness should be instigated amongst stakeholders so as to produce effective interventions at different levels.

Social cohesion and engagement: Civil society

Civil society participants also mentioned having made several inroads in addressing refugees' challenges by coordinating programmes (social

cohesion activities) involving both citizens and refugees to promote cultural tolerance. This was essential in the fight against social exclusion and the associated deprivations because it created what Putman (2000 cited by Cederberg[32]) terms 'bridging forms of social capital' and what Telles[48] terms 'vertical relationships'. Through this form of social capital, disadvantaged groups can overcome their group identities, and go beyond the tribal, racial and cultural boundaries while benefiting from other groups monopolies over opportunities. Another form of social cohesion is one whereby civil society organisations micro financed income-generating projects that were jointly implemented by refugees and South African citizens. We recognise the importance of such activities in deterring social exclusion because social disengagement (a lack of participation in social activities) is one of the key forms through which social exclusion manifests.[43]

A group's ability to harness cultural and symbolic capital is also essential in the deterrence of social exclusion because social capital cannot be examined far from social exclusion (due to its focus on the role of networks and human relationships as assets). A lack of certain forms of capital (cultural and symbolic) can increase refugees' vulnerability, the importance of such programmes thus cannot be undermined due to how they often help improve the level of social contact between migrants and citizens.

The interviews also revealed that civil society was embarking on paradigm shifts aimed at promoting bonding forms of social capital for refugees. Given the existence of top-down power hierarchies within communities (that contribute to the exclusion/disaffiliation of perceived out-groups)[49], the civil society hosting of community workshops (themed along topics of acceptance and tolerance), helped distil some of the group-based exclusionary nodes[50]. A faith-based organisation representative said that theology was also important in inculcating tolerance amongst citizens. On the question of how they alleviate some of the challenge's refugees face in socially integrating, the representative said:

> *…we normally have workshops and we have presenters who talk about the theology of strangers, the theology of migration that all humans are in the image of God regardless of nationality, race, etc. So, if churches can emphasise the message of how all are in the image of God regardless of where you are coming from, let us accept one another – that will help reduce the friction.*

Advocacy and rights sensitisation: Civil society

Issues of human rights education are very important and challenges in effectively sensitising the public on such content is tantamount to a limitation in the implementation of refugee policy. In South Africa, there is a lack of knowledge on such rights not only amongst refugees but also amongst street level bureaucrats.[27] In the furtherance of what Berten and Leisering[4] refer to as the 'bottom-up theorisation' of social protection policy, civil society representatives mentioned how they routinely hold sensitisation workshops with refugees and street level bureaucrats, focusing on refugee rights in the country. Through human rights education and similar interventions,[49] civil society is thus a principal player in the creation of communities where refugee rights are accessible. A civil society participant commented, saying that it is only after workshops that government employees become aware of key issues relating to refugee rights:

> *…workshops should also be done on the rights of refugees nationwide, this is very important because there are a lot of people who do not know about the rights of refugees, you will find that only after a workshop someone will say I did not know.*

The LHR, an implementing partner to the UNHCR, focusing on the legal aspects of refugee lives across South Africa was also interviewed. Although the *Refugee Act 130 of 1998* exempts refugees from some discriminative policies that apply to resident foreign nationals,[27] the South African government has in some instances enacted restrictive measures to disfranchise them. In such instances, the LHR has lobbied against prejudicial laws that disaffiliate refugees and promote deprivations and institutional biases i.e. as exemplified by the Department of Home Affairs systematic closure of refugee reception offices.[44] In response to such office closures (which adversely hamper refugees' access to legal documents), the LHR took the Department of Home Affairs (DHA) to court. An interviewee representing the LHR said that:

> *Our organisation is here to enforce and protect refugee rights in all aspects therefore where we need to litigate, we will litigate against DHA….*

Other instances where similarly exclusionary laws have been successfully challenged through litigations by the civil society include the 2002 declaration on the right to work for asylum seekers and the 2003 legal action against the Director and Minister of Social Development. Owing to such litigations, in 2003 the government availed social protection amenities that are not enshrined within the *Refugee Act 130 of 1998* and these include the South African Social Security Agencies which although not contained in the policy framework, they are in line with global best practices on refugee social protection.[13] The availability of such social safety nets is essential because service exclusion is one of the numerous ways through which deprivations and social closure is perpetuated.[47] A legal attorney from the LHR said:

> *Recognised refugees can also access social grants, which was initially not in the Refugee Act 130 of 1998 but it was a case that was actually brought to the courts that then created the judicial precedence that recognises refugee access to social support grants.*

In facilitating for refugee's inclusion and social protection in South Africa, the LHR also offers migrants free legal assistance through their four legal clinics in Durban, Johannesburg, Pretoria and Mussina. The LHR continuously engages with the UNHCR to bring about awareness on the numerous issues affecting refugees in South Africa.

State and NGO partnerships

Collaboration (dependency relationships)

In exploring what are referred to as dependency relationships,[29] we examined the working relationship between the government and civil society organisations. The United Nations recognises the importance of such cross-sectoral partnerships, with the Comprehensive Refugee Response Framework going beyond the state–civil society partnerships and extending the model to encapsulate think tanks, financial institutions, academia, faith leaders and so forth.[51] While Rosenblum and Post[52] have described civil society as both 'autonomous to the state as well as being dependent on the state', due to how social movements often facilitate the turning of grievances into a form of collective injustice, and then action,[43] in certain instances, they often operate either independently or against the state. An analysis of the level of cordiality between the two sectors in implementing refugee policy was thus carried out in recognition of civil society's role in the global fight against prejudice, inequity and social exclusion.[17] The available studies focusing on refugees in South Africa, the *Refugee Act 130 of 1998* and those focusing on the concept of social exclusion,[5,53] do not discuss such issues within a comparative analytical framework. As contained in this paper, a comparative analytical framework examines the role of the state as well as civil society in alleviating refugees' challenges.

We discovered that civil society in South Africa assumes a relatively collaborative partnership with the state in precluding refugees' social disaffiliation and associated deprivations. This collaboration was best exemplified by the Department of Sport and Recreation's donation of trophies towards some of the social cohesion sporting activities hosted by the civil society in Pietermaritzburg. Khan et al.[43] also identified a multi sectorial approach (integration and collaborations) between the state and civil society as an effective intervention model against social exclusion. To foster information sharing, a participant from a faith-based organisation mentioned how they often invite government department

representatives to their community workshops. In the long run, such dependency relationships improve bureaucratic efficiency on the side of government, as this creates a channel of communication with the refugee communities.[28] Such a line of communication is also essential in whistle-blowing against unruly practices, institutional biases and exclusion within areas of public service.[6]

The prevalence of exploitative structural systems[33,42] can also be mitigated through the establishment of the above-mentioned channels of communication. In reference to such workshops, undertaken with the support of government, a civil society representative said:

> *We are in close relations with government, the Department of Social Development, the SAPS, Department of Sport and Recreation, Department of Health, DHA, Department of Education. Officials are invited from these government departments during workshops.*

Civil society groups also made efforts in discouraging community attitudes that promote inequities and deprivations while at the same time advocating for refugees' social cohesion.[43] In keeping with what Chambers and Kopstein[1] term as being 'in dialogue with the state' and achieving what Weber[41] in his bureaucratic theory refers to as bureaucratic efficiency, the interviews revealed that civil society was facilitating communication between the refugees and local government. Apart from existential monopolizations, other studies have shown that refugees' deprivations in South Africa are also worsened by municipal workers' ignorance of refugee rights and issues.[31,54,55] The social services that municipalities are legally mandated to provide to refugees include trading licences, municipal trading sites, low-income accommodation, among others. During the interviews, the CRA representative mentioned how they were engaged in negotiations with the officials:

> *We always try to be in touch with the officials, but nothing has come up from them. So briefly, there is no channel of communication between refugees and officials in government.*

The ambiguities in the Refugee Act regarding probable channels of communication between refugees and government also debilitate bureaucratic efficiency,[26] as they impede refugees from establishing any form of rapport with the government. Although the duty of civil society is in part to engage the state in dialogue,[1] this has not been the case for the CRA. Such a void in communication has adversely worsened the many challenges refugee status (section 24) permit holders already face in South Africa.[7]

Conclusion

In delimiting the sample (to focus on section 24 permit holders), as well as determining the social protections due to such a group (as enshrined within the *Refugee Act 130 of 1998*), the study was informed by the South African legal framework. From a legal standpoint, access to the rights within the Refugee Act were understood as being contingent on one's possession of a section 24 permit. Ipso facto, the assumptions drawn from the analysis of data were only indicative of the situation with regard to section 24 permit holders (regardless of the conditions that led to the individuals seeking refuge in South Africa, e.g. conflicts, economic issues).

As compared to other African states with refugee camp settlement systems, a free settlement system as found in South Africa is prone to jurisdictional limitations, i.e. within a free settlement system, refugee rights are primarily enshrined within the Constitution. Consequently, 'the government (rather than the 'international community' or NGOs) becomes solely responsible for the social protection of forced migrant groups'.[45] However, if the government does not fully accord such groups with the requisite social protections, they become exposed to a spectrum of risks. This was exemplified by how most refugee participants in the study complained of neglect and the absence of support systems towards social integration and life-skills training.

In the face of the many aforesaid challenges that are often synonymous with free settlement systems, it is of paramount importance to curb the existential gap between the states' formulation of progressive refugee policy frameworks and the policy's dismal implementation (due to credentialism, racial/ethnic monopolisation and social exclusion). This disjuncture was conceptualised in the paper as principally compromising the efficiency of the state (as a bureaucratic institution) in fully operationalising the refugee policy edicts. Multisectoral approaches were then examined, with the determination that the prevailing partnerships between civil society and government in South Africa are multidimensional, i.e. collaborative but mostly conflictive.[52]

Collaborative partnerships were identified as existing between civil society organisations and state departments such as the Department of Social Development, South African Police Service, Department of Home Affairs, Department of Health and Department of Education. These collaborative partnerships (which resulted in a greater degree of bureaucratic efficiency), mainly consisted of information sharing, co-hosting and co-organising of workshops, and social cohesion events. Engagements with the state were also riddled with several challenges – a situation which made collaborative partnerships (dependency relationships) a challenge.

We identified how conflictive partnerships between the two sectors in the country were mostly emanating from a poor working rapport. In the absence of reasonable communication avenues through which effective engagements could be fostered between the two, litigations were the most common and effective methods utilised by civil society in restraining some of the government's excesses. Cases under legal contestation included but were not confined to the nationwide closure of refugee reception offices and the denial of the right to family for forced migrant groups. In advocating for the well-being and integration of refugee populations in South Africa (through mediatory functions, advocacy, human rights education, etc.), it was noted that civil society groups often have to work in autonomy to the state. Although not as extreme as the conflictive type of partnerships, autonomy was also revealed in the paper as presenting several challenges which debilitate the efforts to fully protect refugees. Through the adoption of multisectoral consociations, the state and civil society can forge formidable panaceas against some of the complex (multidimensional) challenges facing refugee groups in the country.

Authors' contributions

S.M. conceived the presented idea and developed the theory while S.R. gave some conceptual guidance. S.M. developed the methodology and analytical methods. S.R. also supervised the data collection and analysis of findings. Both authors discussed the results and contributed to the final manuscript.

References

1. Chambers S, Kopstein J. Bad civil society. Polit Theory. 2001;29(6):837–865.
2. Fischer A. Inequality and the universalistic principle in the post-2015 development agenda [document on the Internet]. c2012 [cited 2019 Dec 12]. Available from: http://www.networkideas.org/wp-content/uploads/2016/09/Development_Agenda.pdf
3. Dongier P, Van Domelen J, Ostrom E, Ryan A, Wakeman W, Bebbington A, et al. Community driven development [document on the Internet]. c2003 [cited 2019 Dec 15]. Available from: http://siteresources.worldbank.org/INTPRS1/Resources/383606-1205334112622/5805_chap9.pdf
4. Berten J, Leisering L. Social policy by numbers. How international organisations construct global policy proposals. Int J Soc Welf. 2016;26(2):151–167.
5. Spreen C, Vally S. Education rights, education policies and inequality in South Africa. Int J Educ Dev. 2006;26(4):352–362. https://doi.org/10.1016/j.ijedudev.2005.09.004
6. Masuku S. The implementation of the Refugee Act 130 of 1998 in South Africa and the question of the social exclusion of forced migrants: A case study of DRC forced migrants in Pietermaritzburg [thesis]. Pietermaritzburg: University of KwaZulu Natal; 2018.
7. Crush J, Ramachandran S, Pendleton W. Soft targets: Xenophobia, public violence and changing attitudes to migrants in South Africa after May 2008. Migration Policy Series no. 64. Cape Town: Southern African Migration Programme; 2013 [cited 2019 Dec 12]. Available from: https://www.africaportal.org/publications/soft-targets-xenophobia-public-violence-and-changing-attitudes-to-migrants-in-south-africa-after-may-2008/

8. Valenta M, Strabac Z. State-assisted integration, but not for all: Norwegian welfare services and labour migration from the new EU member states. Int Soc Work. 2011;54(5):663–680. https://doi.org/10.1177/0020872810392811

9. White H. Asylum in tension: Germany and the European migrant crisis [master's thesis]. Auckland: University of Auckland; 2017. http://hdl.handle.net/2292/32926

10. Gerber P. Education about human rights. Altern Law J. 2011;36(4):245–249. https://doi.org/10.1177/1037969x1103600406

11. Owusu S, Kanyane B, Viljoen J, Tshitiza O, Wentzel M, Sakoane L. A critical skills attraction index for South Africa: Final fieldwork report [document on the Internet]. c2016 [cited 2019 Dec 12]. Available from: http://hdl.handle.net/20.500.11910/10889

12. Crush J. South Africa: Policy in the face of xenophobia [webpage on the Internet]. c2008 [cited 2019 Dec 12]. Available from: http://www.migrationpolicy.org/article/south-africa-policy-face-xenophobia/

13. United Nations. Universal declaration of human rights [webpage on the Internet]. c1951 [cited 12 December 2019]. Available from: https://www.un.org/en/universal-declaration-human-rights/

14. United Nations. The 1951 Refugee Convention [webpage on the Internet]. c1951 [cited 2019 Dec 12]. Available from: http://www.unhcr.org/1951-refugee-convention.html

15. Murphy R. Social closure. The theory of monopolization and exclusion. Br J Sociol. 1989;30(1):171.

16. Hebbani A, Colic-Peisker V. Communicating one's way to employment: A case study of African settlers in Brisbane, Australia. J Intercult Stud. 2012;33(5):529–547. https://doi.org/10.1080/07256868.2012.701609

17. Harbeson J, Rothchild D, Chazan N. Civil society and the state in Africa. London: Lynne Rienner Publishers; 1994.

18. Shahin Y, Kapur A, Seita A. Diabetes care in refugee camps: The experience of UNRWA. Diabetes Res Clin Pract. 2015;108(1):1–6. https://doi.org/10.1016/j.diabres.2015.01.035

19. Chelwa N, Likwa R, Banda J. Under-five mortality among displaced populations in Meheba refugee camp, Zambia, 2008–2014 [webpage on the Internet]. c2016 [cited 2019 Dec 12]. Available from: https://archpublichealth.biomedcentral.com/articles/10.1186/s13690-016-0161-9

20. Kairu P. East Africa host countries at a crossroads: Are refugees welcome or not? – Uganda [webpage on the Internet]. c2018 [cited 2019 Dec 12]. Available from: https://reliefweb.int/report/uganda/east-africa-host-countries-crossroads-are-refugees-welcome-or-not

21. Parkin F. The social analysis of class structure. London: Tavistock Publications; 1974.

22. Fukuyama F. Social capital and development: The coming agenda. SAIS Rev. 2002;22(1):23–37. https://doi.org/10.1353/sais.2002.0009

23. Fukuyama F. Social capital, civil society and development. Third World Q. 2001;22(1):7–20. https://doi.org/10.1080/713701144

24. Weber M. Economy and society. London: University of California Press; 1979.

25. Selznick P. An approach to a theory of bureaucracy. Am Sociol Rev. 1943;8(1):47–54. https://doi.org/10.2307/2085448

26. Brauns M. Public healthcare in a post-apartheid South Africa: A critical analysis in governance practices [dissertation]. Pietermaritzburg: University of KwaZulu-Natal; 2016.

27. Fraser N. Unruly practices. Minneapolis, MN: University of Minnesota Press; 1989.

28. Murphy R. Weberian Closure Theory: A contribution to the ongoing assessment. Br J Sociol. 1986;37(1):21–41. https://doi.org/10.2307/591049

29. Hill M, Hupe P. Implementing public policy. 3rd ed. London: Sage; 2014.

30. Weber M. Economy and society: An outline of interpretive sociology. Berkeley, CA: University of California Press; 1978.

31. Morrow RA. Review of 'Social closure: The theory of monopolization and exclusion'. Can J Sociol. 1990;15(4):477–482. https://doi.org/10.2307/3341133

32. Cederberg M. Migrant networks and beyond: Exploring the value of the notion of social capital for making sense of ethnic inequalities. Acta Sociol. 2012;55(1):59–72. https://doi.org/10.1177/0001699311427746

33. Galtung J. Violence, peace, and peace research. J Peace Res. 1969;6(3):167–191. https://doi.org/10.1177/002234336900600301

34. Kabeer N. Social exclusion, poverty and discrimination towards an analytical framework. IDS Bull. 2000;31(4):83–97. https://doi.org/10.1111/j.1759-5436.2000.mp31004009.x

35. Denzin N, Lincoln Y. The landscape of qualitative research. 1st ed. Thousand Oaks, CA: Sage; 2003.

36. Denzin N, Lincoln Y. The landscape of qualitative research. 2nd ed. Los Angeles, CA: Sage; 2008.

37. Hyett N, Kenny A, Dickson-Swift V. Methodology or method? A critical review of qualitative case study reports. Int J Qual Stud Health Well-being. 2014;9(1), Art. #23606, 12 pages. https://doi.org/10.3402/qhw.v9.23606

38. Gray D. Doing research in the real world. 1st ed. London: Sage; 2013.

39. Kelman HC. Ethical issues in different social science methods. In: Beauchamp TL, Faden RR, Wallace RJ, Walters L, editors. Ethical issues in social science research. Baltimore, MD: Johns Hopkins University Press; 1982. p. 40–98.

40. Etikan I, Musa SA, Alkassim RS. Comparison of convenience sampling and purposive sampling. Am J Theor Appl Stat. 2016;5(1):1–4. https://doi.org/10.11648/j.ajtas.20160501.11

41. Thorne SE. Review of 'Max Weber: The theory of social and economic organization'. Yale Law J. 1948;57(4):676–678. https://doi.org/10.2307/793128

42. Mosse D. Power and the durability of poverty: A critical exploration of the links between culture, marginality and chronic poverty. Chronic Poverty Research Centre working paper no. 107 [document on the Internet]. c2007 [cited 2020 Jan 28]. http://dx.doi.org/10.2139/ssrn.1615629

43. Khan S, Combaz E, McAslan Fraser E. Social exclusion: Topic guide. Birmingham, UK: Governance and Social Development Resource Centre, University of Birmingham; 2015.

44. Kavuro C. Refugees and asylum seekers: Barriers to accessing South Africa's labour market. Law Dem Dev. 2016;19(1):232–260. https://doi.org/10.4314/ldd.v19i1.12

45. Hungwe C. Surviving social exclusion: Zimbabwean migrants in Johannesburg, South Africa [dissertation]. Pretoria: University of South Africa; 2013.

46. Makhema M. Social protection for refugees and asylum seekers in the Southern Africa Development Community (SADC) [document on the Internet]. c2009 [cited 2019 Dec 13]. Available from: http://documents.worldbank.org/curated/en/799011468104641822/pdf/491680NWP0Asyl10Box338943B01PUBLIC1.pdf

47. Amit R. No way in. Johannesburg: African Centre for Migration and Society; 2012. p. 37–73. Available from: http://www.migration.org.za/wp-content/uploads/2017/08/No-Way-In-Barriers-to-Access-Service-and-Administrative-Justice-at-South-Africa%E2%80%99s-Refugee-Reception-Offices.-pdf.pdf

48. Telles E. Race in another America: The significance of skin colour in Brazil. Princeton, NJ: Princeton University; 2004. https://doi.org/10.1515/9781400837434

49. Leggett T. Still marginal: Crime in the coloured community. S Afr Crime Quart. 2004;7:21–26. https://doi.org/10.17159/2413-3108/2004/v0i7a1047

50. Fozdar F. Social cohesion and skilled Muslim refugees in Australia. Aust N Z J Sociol. 2011;48(2):167–186. https://doi.org/10.1177/1440783311413482

51. United Nations. Comprehensive Refugee Response Framework: Delivering more comprehensive and predictable responses for refugees [webpage on the Internet]. c2018 [cited 2019 Dec 13]. Available from: https://www.unhcr.org/comprehensive-refugee-response-framework-crrf.html

52. Rosenblum N, Post R. Civil society and government. Princeton, NJ: Princeton University Press; 2002.

53. Manicom D, Mullagee F. The status of asylum seekers and refugees in South Africa: An independent overview. Afr Insight. 2010;39(4), Art. #54672. https://doi.org/10.4314/ai.v39i4.54672

54. Adato M, Carter M, May J. Exploring poverty traps and social exclusion in South Africa using qualitative and quantitative data. J Dev Stud. 2006;42(2):226–247. https://doi.org/10.1080/00220380500405345

55. Palmary I. Refugees, safety and xenophobia in South African cities: The role of local government [document on the Internet]. c2002 [cited 2019 Dec 13]. Available from: https://www.csvr.org.za/docs/foreigners/refugeessafteyand.pdf

AUTHORS:
Eduan Kotzé[1]
Burgert A. Senekal[2]
Walter Daelemans[3]

AFFILIATIONS:
[1]Department of Computer Science and Informatics, University of the Free State, Bloemfontein, South Africa
[2]Department of South African Sign Language and Deaf Studies, University of the Free State, Bloemfontein, South Africa
[3]CLiPS Research Centre, University of Antwerp, Antwerpen, Belgium

CORRESPONDENCE TO:
Eduan Kotzé

EMAIL:
kotzeje@ufs.ac.za

DATES:
Received: 28 Aug. 2019
Revised: 22 Oct. 2019
Accepted: 01 Nov. 2019
Published: 26 Mar. 2020

HOW TO CITE:
Kotzé E, Senekal BA, Daelemans W. Automatic classification of social media reports on violent incidents in South Africa using machine learning. S Afr J Sci. 2020;116(3/4), Art. #6557, 8 pages. https://doi.org/10.17159/sajs.2020/6557

ARTICLE INCLUDES:
☒ Peer review
☐ Supplementary material

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

EDITOR:
Tania Douglas

KEYWORDS:
WhatsApp, text classification, Word2Vec, protests, open-source intelligence, OSINT

FUNDING:
None

Automatic classification of social media reports on violent incidents in South Africa using machine learning

With the growing amount of data available in the digital age, it has become increasingly important to use automated methods to extract useful information from data. One such application is the extraction of events from news sources for the purpose of a quantitative analysis that does not rely on someone needing to read through thousands of news articles. Overseas, projects such as the Integrated Crisis Early Warning System (ICEWS) monitor news stories and extract events using automated coding. However, not all violent events are reported in the news, and while monitoring only news agencies is sufficient for projects such as ICEWS which have a global focus, more news sources are required when assessing a local situation. We used WhatsApp as a news source to identify the occurrence of violent incidents in South Africa. Using machine learning, we have shown how violent incidents can be coded and recorded, allowing for a local level recording of these events over time. Our experimental results show good performance on both training and testing data sets using a logistic regression classifier with unigrams and Word2vec feature models. Future work will evaluate the inclusion of pre-trained word embedding for both Afrikaans and English words to improve the performance of the machine learning classifier.

Significance:

- The logistic regression classifier using TFIDF unigram, CBOW and skip-gram Word2Vec models were successfully implemented to automatically analyse and classify WhatsApp messages from groups that share information on protests and mass violence in South Africa. At the time of publishing, messages were collected from 26 WhatsApp groups across South Africa and automatically classified on an hourly basis.

Introduction

Social media has evolved rapidly during the past few years and has become an increasingly popular platform for acquiring opinions and information about events.[1] Popular social media platforms include online social networks, microblogs, wikis, media sharing and reviews.[2] Text mining is often used to mine information from the natural language text generated in these platforms.[3] A functional text analytic application would typically include natural language processing, information retrieval and machine learning, together with linguistics and statistical techniques to extract information from text.[3]

One text mining technique is text classification, which is often considered one of the fundamental tasks in natural language processing. In text classification, supervised machine learning is used to assign a label or probability value to an instance (i.e. sentence or text document). However, other variations of text classification also allow the assignment of multiple labels to an instance. These labels could be continuous values, but, generally, the classification problem assumes categorical or binary (i.e. 0 or 1) values for the labels.[4] For the purpose of this study, we focus on text classification as the main text mining technique. Our objective is to gain insight into the linguistic aspects of reports about violent incidents in South Africa by collecting, annotating and classifying a user-generated data set. We explore the text characteristics (features) that are potentially useful in distinguishing between events and non-events, and apply these features in several machine-learning algorithms.

Text classification

Text mining applications play an important role in knowledge discovery as the volume and diversity of digital texts have increased significantly.[5] Popular text mining applications include email classification, spam filtering, categorising news articles or autotagging customer queries, sentiment analysis and opinion mining.[6,7] These examples are known as text classification applications and involve the assignment of labels to data (called annotation), using this labelled data as training data for supervised learning of a classifier, and then using this classifier to label (or predict) unlabelled documents.[8,9] The focus of any text classification application is to classify a single object into a discrete class by extracting useful features (attributes) from the object. Features – or the set of attributes, represented as a vector – could for example be words extracted from the text. A classification model typically relates the features in the underlying record to one of the class labels. Machine-learning algorithms are used to construct a classification model and predict the class label (binary, multiclass, multilabel) of a new text document. For this study, we made use of a supervised multi-class machine-learning classifier in which a document is assigned to one class among several possible classes.

Related research

The automatic classification of text is a well-known natural language processing research problem for which there has been an extensive history of scholarly work.[7] However, in the context of defence and security, more work is required to automatically classify news reports related to emergency information. Several initiatives to extract information automatically from news sources have been launched overseas. Examples include the Policon model[10-12], which has been used by the US Central Intelligence Agency (CIA)[13], and Senturion[14], which has been used by the US Department of Defense[13]. Lockheed Martin's Integrated Crisis Early Warning System extracts

structured data from news.[13] Other similar projects include the CIA's Political Instability Task Force[15-17], the Armed Conflict Location and Event Data project[18], the Uppsala Conflict Data Program[19,20], and the Social Conflict Analysis Database (SCAD)[21]. See Schrodt[22] for a discussion of similar projects.

Our project differs from these projects in a number of ways. In those projects, not all events are shared by the main news agencies, as limited space and resources compel them to focus on larger events.[23,24] We required a more comprehensive data set as our outlook is local and not global. In addition, people no longer rely solely on mainstream media to access information. We therefore turned to WhatsApp as a news source. WhatsApp is a free messaging service owned by Facebook which enables users to message, call and share files. WhatsApp has become increasingly popular as a source of news[25]; the use of WhatsApp for news has almost tripled since 2014 in many countries.[26] In South Africa, WhatsApp is also used regularly to share emergency information.[27]

Data collection and annotation

Data collection

Messages were collected from a variety of WhatsApp groups (group chat is a feature in which the same message can be sent at the same time to multiple contacts who opt in), which cannot be named for ethical reasons. Most operate within the Afrikaans and English communities (urban and rural) across South Africa and focus on protests and serious crimes, such as hijackings, cash-in-transit robberies and farm attacks. The majority of messages are eyewitness accounts, and while a much smaller number come from mainstream news agencies, these cover a wide spectrum of the South African media. Numerous messages also come from Facebook groups and from Twitter. It should be noted that the groups were selected based on their activity and dedication to accuracy. None of these groups deliberately spread false news and while an incident may later be shown to have been false, our objective is not to establish crime statistics but rather to classify the messages themselves.

Because WhatsApp is a closed platform, we first had to obtain ethical clearance and permission to be added to WhatsApp groups that share information on violent incidents. Ethnical clearance was granted for this project by the General/Human Research Ethics Committee of the University of the Free State (UFS-HSD2019/0175).

Data were collected from 15 WhatsApp groups between 30 May 2018 and 18 February 2019. In total, 23 360 WhatsApp messages were retrieved in either English or Afrikaans. After filtering the corpus for and removing duplicate messages, the experimental corpus comprised 8398 unique messages.

Data preparation

Text pre-processing is required because online social media data tend to be very 'noisy' and contain uninformative parts, such as HTML tags.[28] We followed the basic linguistic processing steps required to prepare the lexical source for text classification. This processing included data cleaning, removal of stopwords, tokenisation, as well as syntactic parsing.[29] One of the first steps was message and sentence tokenisation. All the messages in the WhatsApp data set were stored as either single or multiline messages. We separated each multiline message into several single-line messages. Using regular expressions, we identified multiline messages and converted them into the following single-line message format: [date, time] + [From]: [Message]. UTF-8 file encoding was performed during the file reading operation to avoid character set conversion errors. Next, all punctuation marks and special characters, including emoticons, were removed from each text message. HTML tags such as web addresses ('http:' and 'www') were also stripped from the text message, followed by white space removal and text lowercasing. Text messages that contained contractions, slang words and hyphenated words were transformed into regular expressions using a custom dictionary (for example, the Afrikaans 'ek't' was converted to 'ek het', and the English 'isn't' to 'is not'; 'Potch' to 'Potchefstroom' and 'CT' to 'Cape Town'). The custom dictionary contained all possible contractions in Afrikaans and English. Misspelled words were identified and normalised, again making use of a custom dictionary. Hyphenated words were reconstructed as hyphens were considered a special character and were removed earlier. We also made use of a lookup dictionary to reconstruct hyphenated words (for example, 'hi jacking' became 'hi-jacking'). All custom and lookup dictionaries were manually constructed using a trial-and-error approach as we identified contracted, misspelled and hyphenated words. Due to the bilingual nature of the corpus, stemming and lemmatisation were not performed. Finally, all stopwords were removed from the corpus. For English words, we made use of the stopword list in NLTK[30] and for Afrikaans words, we translated the English stopwords into their corresponding Afrikaans words.

Data annotation

Two annotators were asked to identify whether the text under investigation was related to an event. If the message was considered to be an event, annotators identified the type of event. If not, it was considered a 'safe' message. The four event-related categories identified were: farm attack, land grab, crime and protest. An example of each category is given in Table 1.

Annotation statistics

All the messages were independently annotated for event detection by two annotators, A and B, who were postgraduate students familiar with linguistics and the South African environment, and are English second-language speakers. When A and B did not agree, a third annotator functioned as a tiebreaker to obtain a gold standard. The inter-annotator agreement between the two annotators was calculated using Cohen's kappa.[31] The agreement score was $\kappa=81.72$, which indicated substantial agreement. Of the 8395 unique messages, both annotators agreed on

Table 1: Examples of categories related to events

Annotation category	Definition	Annotation example [sic]
Farm attack	A violent attack that occurs on a farm or smallholding, including assault, murder, attempted murder, robbery or rape	FARM ATTACK!!!!! Farmattack last night around 03:00 at smallholding / farm in the Tarlton area next to Krugersdorp. According to information received, both the man and the woman, in their 70's, were beaten during the attack. The man is also bedridden. Luckily no bone fractures were sustained. The man was also burned with boiling water. They put the iron on to burn the woman, but something disturbed them and they fled. They both were tied up with ropes after the attack and the woman was only able to free herself after three hours. She then climbed through the window to get some help from neighbors. Names to be withheld until police release more information. The attackers left with firearms and money. No arrests has been made.
Land grab	An illegal occupation of land with the intent to settle	VIRGINIA_ (17h20) Unrest reported on the R73 towards Winburg (14h00) Land grab reported on a farm in the area.
Crime	Other forms of crime that are not farm attacks, land grabs or illegal protests, e.g. robbery, theft and murder in cities or towns	TUGELA TOLL PLAZA_ (16H30) Cash in Transit Robbery reported.
Protest	A violent or potentially violent mass protest	*CAPE TOWN - VIOLENT* *PROTEST - ROAD* *CLOSED* Landsdown Road closed between Philipi and Gugulethu due to Violent protest action - SAPS on scene.

7480 messages. The difference set (n=915) was annotated by two additional annotators, C and D. We implemented a majority-vote algorithm to select annotations in which C and D agreed with either A or B. In this way, 660 annotations were added to the 7480 messages. Further data exploration was conducted to refine the combined corpus (n=8085). All messages shorter than 5 characters or longer than 1000 characters were removed from the corpus. We found that messages that fell outside these length thresholds were mainly informational messages (i.e. 'safe'). The final pre-processed and tokenised data set consisted of 7889 unique messages. After pre-processing the corpus, the data set was split (80/20) into training (n=6311) and testing (n=1578) data sets. The number and classification of each message are shown in Table 2.

Table 2: Data set attributes

	Labelled message	Proportion	Training	Testing
Land grab	50	0.63%	38	12
Farm attack	271	3.43%	223	48
Crime	431	5.46%	352	79
Protest	2949	37.38%	2363	586
Safe	4188	53.09%	3335	853
Total	**7889**	**100.00%**	**6311**	**1578**

Experimental set-up

Feature extraction

Text data are known for their high dimensionality of text features.[4] As a result, feature extraction and feature selection are necessary tasks before any text classification can be attempted. For this study, we made use of two popular feature extraction techniques: bag-of-n-gram (n-grams) and term frequency – inverse document frequency (TF-IDF). In n-grams, the presentation model creates sequences of n words instead of frequency counts of single words (bag-of-words).[32] The features in n-grams are representations of all combinations of adjacent words of length n in the text. In other words, 2-gram or bigrams will represent a model with combinations of two adjacent words in the text document, while 3-gram or trigrams will be combinations of three adjacent words in a text document. In TF-IDF, weights are created instead of frequency counts. The idea is that terms that frequently occur in a document relative to the number of times they occur in the entire corpus, are more important than terms that occur commonly. Both n-gram and TF-IDF models introduce several shortcomings, such as data sparsity and high dimensionality.[33] These phenomena are referred to as 'the curse of dimensionality', which means that, as the feature space increases in dimensions (i.e. character or word tokens), the data become more sparse and less informative to the underlying decision space.[34]

One solution to overcoming these shortcomings is to make use of more recent models in which word representations are based on large natural language corpora by tracking the contexts in which they occur – known as 'word embeddings'. Word embedding is a language modelling technique used to create a continuous higher dimensional vector space to overcome the curse of dimensionality and represent word meaning.[35] A popular language model to produce such word embeddings is Word2Vec.[36] Word2Vec is a group of two-layer neural network models that are trained on a large corpus of text to reconstruct the linguistic context of words. Two architectural Word2vec models available to learn the distributed representation of words are continuous-bag-of-words (CBOW) and skip-gram.[36] In both these models, techniques are used to learn word weights, which act as word vector representations. The CBOW model predicts the word from a window of surrounding words, while the skip-gram model predicts the context words, given the current word.

Considering this background, we took extra care in constructing the data pipeline for the text classification process and, as far as possible, discarded non-relevant words without influencing the text classification outcome. We opted for TF-IDF instead of frequency counts (bag-of-words) because TF-IDF considers the inverse document frequency of each term when performing raw term frequency computations, and thus often produces better results.[37] The following feature types were extracted from both data sets:

- **Word *n*-gram TF-IDF**: features indicating the frequently occurring unigrams (1,1) or bigrams (2,2), or unigrams and bigrams (1,2)
- **Word2Vec**: words to vectors using continuous-bag-of-words (CBOW) and skip-gram models. A function was added to calculate an average of all vectors per message. The function summed all the word vectors occurring in a message and then divided it by the count of the word vectors.

For the first experiment, we extracted TF-IDF features on the training data set using scikit-learn, a machine-learning library for Python.[38] Initially, both the minimum (min_df) and maximum (max_df) document/corpus frequencies were set to '1' for all n-gram models. The maximum number of features (top n max_features) was set to 'None' due to the small size of the corpus. Inverse-document-frequency reweighting was enabled (use_idf=True) and, to reduce document length bias, we set norm to 'L2'. The vectoriser created a sparse matrix and the sparsity metric was calculated for both matrixes. The sparsity metric was calculated by dividing the number of non-zero values by the sum of the number of samples and number of features. The lower the value, the sparser are the data in the matrix, making it more challenging to model. The results are given in Table 3.

Table 3: The TF-IDF n-gram frequencies

n-gram	Number of rows	Number of features	Number of non-zeros	Sparsity
unigram (1,1)	7889	16 926	121 246	0.09%
bigram (2,2)	7889	79 550	124 091	0.02%
unigram and bigram (1,2)	7889	96 476	245 337	0.03%

For feature selection, we used the chi-squared statistic (X^2) to find the optimal number of features. We first calculated the estimated mutual information value for a discrete target variable by measuring the dependency between a feature and the target variable. Some of the highest dependency words included unrest (0.09620), farm (0.05994), plaasaanval (0.05105), attack (0.03793), farm attack (0.03070), robbery (0.03006), land (0.02914), invasion (0.01117) and land invasion (0.00859). Thereafter, we used the chi-squared statistic to find the optimal number of features with the highest value. The top 12 features with higher values, i.e. higher usefulness, are depicted in Figure 1.

We also examined whether top features would have a significant influence on model accuracy. We used the scikit-learn feature_selection module with the SelectPercentile method which, when given a statistical test, selects a percentile of features with the highest score. We used a decision tree classifier and set the function to measure the quality of a split to 'entropy' (for the information gain) and used the chi-squared statistic for feature selection. Feature selection was repeated for unigrams, bigrams and unigrams with bigrams using cross-validation accuracy as a performance metric on the test set. We found no significant difference in accuracy, and opted to use all unigram and bigram features for our experiments as the number of features was manageable. The chi-squared statistic results are given in Table 4.

Table 4: X^2 (chi-squared) statistic results

n-gram	Features	Optimal number of features	Accuracy (top percentile)	Accuracy (all features)
unigram (1,1)	15 110	12 239	0.84981	0.84728
bigram (2,2)	66 289	13 921	0.79278	0.79341
unigram (1,1) and bigram (1,2)	81 399	65 933	0.84411	0.84791

For the second experiment, we created word embeddings with Word2vec on the training data set using both CBOW (sg=0) and skip-gram (sg=1) models. The context window size value was set to 6 and the number of

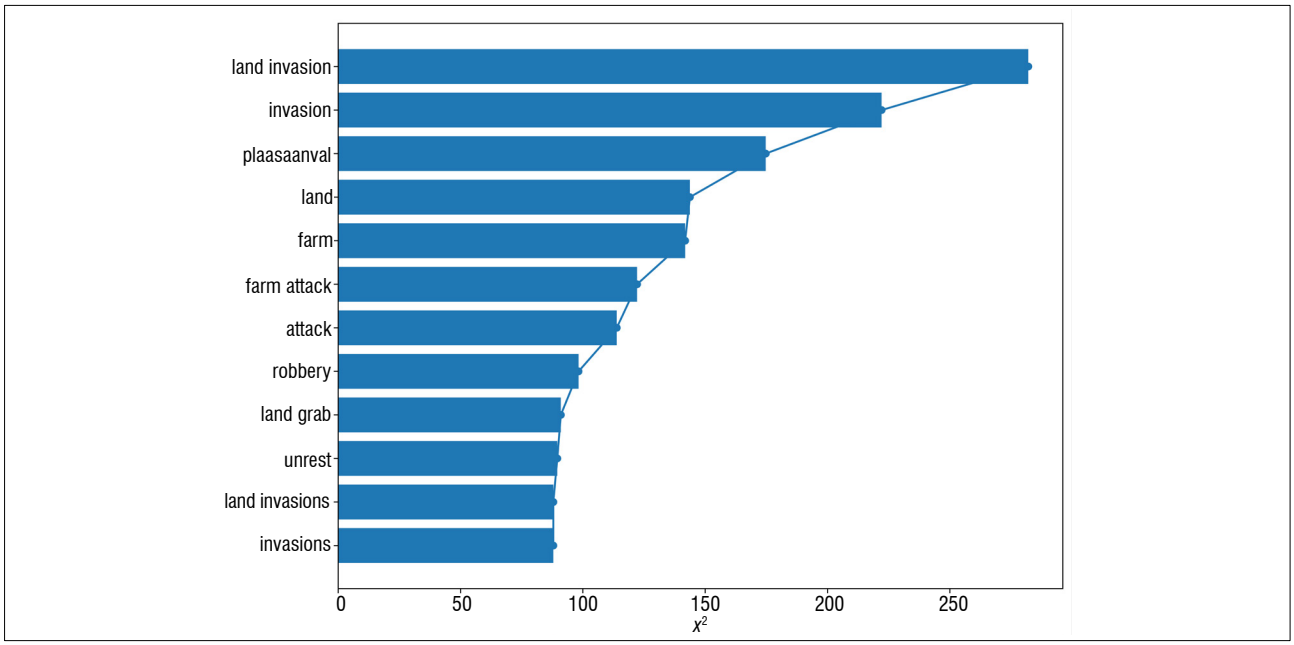


Figure 1: Top X^2 (chi-squared) features.

features (dimension) to 100. Due to the small corpus of messages, we had to set words with an occurrence count as low as possible (min_count=1). Words below this threshold were not kept in the vocabulary. Because each WhatsApp message can have a different number of vectors depending on the number of words it contains, a function was added to calculate an average of all vectors per message. We also checked whether the words in a message occur in the vocabulary of the Word2vec model. If the word was not found, the function returned a 0.0, otherwise the average of the word vectors was returned. Both CBOW and skip-gram models produced a vocabulary of 15 114 words from the training corpus.

Data sampling

The experimental corpus features had a heavy class imbalance, distributed as shown in Table 4, with the large majority of posts not being events (i.e. 'safe'). In machine-learning classification, class imbalance can lead to decreased performance and accuracy.[39] We were also concerned that the machine-learning algorithms would be biased towards the majority class of no incidents ('safe') and treat the minority classes ('Land Grab', 'Farm Attack', and 'Crime') as outliers and ignore the observations. We applied the Python package imbalanced-learn as data resampling to counter this possibility.[40] The methods available included 'over-sampling' and 'under-sampling'. Over-sampling generates new samples in the classes that are under-represented, while under-sampling reduces the number of samples in the targeted classes. We opted for random over-sampling (RandomOverSampler) because of the relatively small number of instances. This method generates new samples in the minority classes ('Land Grab', 'Farm Attack' and 'Crime') by picking samples at random with replacement.

Learning algorithms

From the text classification literature, support vector machines (SVMs), naïve Bayes, decision trees, logistic regression and random forests are popular machine-learning algorithms used to classify text. For the automatic detection of events, we performed text classification experiments using a linear kernel SVM, random forest, gradient boosting machines and logistic regression.

Support vector machines

Support vector machines is a set of versatile supervised learning methods with different kernel functions (linear, polynomial, radial basis function and sigmoid) that can be specified for the decision function. The purpose of a linear kernel classifier is to find a maximum margin hyperplane that can classify data points appropriately. In SVM, we optimised a margin, which is defined as the distance between the separating hyperplane (decision boundary) and the training sample closest to this hyperplane.[41] For our experiments, we used the SVM classier with linear kernel (LinearSVC) implemented in LIBSVM.[42] The advantage of using LIBSVM over the native Python implementation is that it allows for quick training of large numbers of linear classifiers, as long as the data set fits into computer memory.[41]

Random forest

Random forest is an ensemble decision-tree algorithm that involves building large numbers of decision trees and then ensembling their outputs.[43] Because a random forest is meta-estimated, it fits a number of decision tree classifiers on various sub-samples of the data set and uses averaging to improve accuracy. The function to measure the quality of a split was set to 'gini', which is the default, and used for Gini impurity.

Gradient boosting machines

Gradient boosting machines is a much newer machine-learning technique, and, much like a random forest, it is based on ensembling decision trees. However, it uses gradient boosting to iteratively train new models that address the weak points of previous models.[43] The focus of boosting is more to reduce bias than variance, and as a result, boosting tends to improve upon its base models when they have high bias and low variance.[44] When comparing the two ensemble methods, random forest trains each tree independently, using a random sample of the data, whereas gradient boosting machines trains one tree at a time, where each new tree is used to correct errors made by a previously trained tree.

Logistic regression

Logistic regression and random forests are text-processing models when using feature-engineering tools such as bag-of-words or n-grams.[43] The algorithm used in the optimisation problem (solver) was set to 'newton-cg' as it was a multiclass classification problem. Additionally, multiclass was set to 'multinomial', meaning the loss minimised is the multinomial loss fit across the entire probability distribution. As the 'newton-cg' solver only supports L2 regularisation with primal formulation, the value of the penalisation parameter was set to the default value ('L2').

Model selection

The four classifiers were evaluated as multiclass classifiers (classifying instances into three of more classes). Table 5 shows the hyperparameter combinations for model selection.

Table 5: Hyperparameters in grid-search model selection

Algorithm	Hyperparameter	Values	Optimal Value
Support Vector Machine	Penalty parameter C of the error term (C)	$1e^{\{-3,-2,-1,0,1,2\}}$	1
	Tolerance for stopping criteria (tol)	$1e^{\{-3,-2,-1,0,1\}}$	0.001
RandomForestClassifier	Number of trees in the forest (Estimators)	[50,100,150,200]	200
GradientBoosting	The number of boosting stages to perform (Estimators)	[50,100,150,200]	200
	Shrinks the contribution of each tree by learning_rate	[1, 0.5, 0.25, 0.1, 0.05, 0.01]	0.25
LogisticRegression	Inverse of regularisation strength (C)	$1e^{\{-3,-2,-1,0,1,2\}}$	100
	Tolerance for stopping criteria (tol)	$1e^{\{-3,-2,-1,0,1\}}$	10

Model selection was done using 10-fold cross-validation in an exhaustive grid search over all possible hyperparameter configurations. The grid search followed by a 10-fold cross-validation was performed for both experiments (TFIDF and TDIDF + Word2vec) using the training data set.

Evaluation measures

Precision, recall, F1-score and accuracy are evaluation metrics used to assess the performance of algorithms in text classification. As we were using a multiclass classifier, calculating the performance metrics was more complicated than for a binary classifier (i.e. only two labels). Measures for a multiclass classifier are based on the generalisation of the binary metrics (described above) making use of either macro-averaging or micro-averaging. Macro-averaging calculates the mean of the binary metrics, while micro-averaging gives each sample-class pair an equal contribution to the overall metric.[45] For this study, we used M and μ indices to represent the metrics in macro-averaging and micro-averaging. In macro-averaging, precision_M is defined as the average per-class agreement of the data class labels with those of the classifier. Recall_M (or sensitivity) is the average per-class effectiveness of a classifier to identify class labels. F-score_M is the relationship between the sample's positive labels and those predicted by the classifier on a per-class average.

$$Precision_m = \frac{\sum_{i=1}^{l}\left(\frac{tp_i}{tp_i+fp_i}\right)}{l} \quad \text{Equation 1}$$

$$Recall_m = \frac{\sum_{i=1}^{l}\left(\frac{tp_i}{tp_i+fn_i}\right)}{l} \quad \text{Equation 2}$$

$$F-score_m = \frac{(\beta^2+1)Precision_m Recall_m}{\beta^2 Precision_m Recall_m} \quad \text{Equation 3}$$

$$Average\ Accuracy = \frac{\sum_{i=1}^{l}\left(\frac{tp_i+tn_i}{tp_i+fn_i+fp_i+tn_i}\right)}{l} \quad \text{Equation 4}$$

where tp is true positive, tn is true negative, fp is false positive and fn is false negative.

For micro-averaging, we report only the F-score_μ because micro-averaging in a multiclass setting produces precision_μ, recall_μ and F-score_μ that are all identical to accuracy.[46]

Experiments and results

We discuss the performance of the different models on the training data set, obtained in 10-fold cross-validation, and on the test data set. The accuracy performance evaluation parameter was calculated and used to determine the best performance. We report the accuracy, precision, recall and F-scores for the five event classes.

Results on the training data

We tested the performance of the four models on the training data using 10-fold cross-validation. When the data set was created it was unbalanced, but as part of the pipeline, random over-sampling was used to address the uneven distribution of classes (see 'Data sampling'). A detailed comparison of all scores can be found in Table 6. The word unigram model obtained the highest score for each model, with the highest training F-score_M of 0.736 (s.d.=0.11) for the logistic regression classifier.

Results on the test set

We also tested the performance of the four models on the test set, the results of which can be found in Table 7. On the test set, all the word unigram models reached F-scores_M of 0.647 or higher compared to training F-scores_M of 0.580. The reason for a slightly higher performance of the n-gram models compared to the training set results could be sparsity. Because we have a small number of instances, it is possible that some of the unigrams or bigrams in the training set might not have occurred in the testing set. We also included a majority baseline to help contextualise the results. We used a DummyClassifier[38] with a strategy set to most_frequent which always predicts the most frequent

Table 6: Results on the training data set

Features	Model	Acc	R_M	P_M	$F1_M$	$F1_\mu$ †
word unigram (1,1)	Support vector machine	0.884 (±0.02)	0.684 (±0.11)	0.770 (±0.19)	0.711 (±0.13)	0.884 (±0.02)
	Random forest	0.864 (±0.01)	0.542 (±0.06)	0.689 (±0.16)	0.580 (±0.07)	0.864 (±0.01)
	Gradient boosting	0.868 (±0.02)	0.711 (±0.10)	0.675 (±0.08)	0.684 (±0.07)	0.868 (±0.02)
	Logistic regression	**0.886** (±0.02)	**0.721** (±0.09)	**0.775** (±0.15)	**0.736** (±0.11)	**0.886 (±0.02)**
word bigram (2,2)	Support vector machine	0.831 (±0.02)	0.518 (±0.09)	**0.730** (±0.21)	0.569 (±0.12)	0.831 (±0.02)
	Random forest	0.787 (±0.02)	0.426 (±0.03)	0.638 (±0.06)	0.467 (±0.04)	0.787 (±0.02)
	Gradient boosting	0.792 (±0.03)	**0.551** (±0.13)	0.692 (±0.17)	**0.590** (±0.12)	0.792 (±0.03)
	Logistic regression	**0.833** (±0.03)	0.538 (±0.09)	0.724 (±0.20)	0.589 (±0.12)	**0.833 (±0.03)**
word unigram (1,1) and word bigram (1,2)	Support vector machine	**0.883** (±0.02)	0.679 (±0.11)	0.767 (±0.19)	0.703 (±0.13)	**0.883 (±0.02)**
	Random forest	0.852 (±0.02)	0.504 (±0.07)	0.713 (±0.17)	0.550 (±0.09)	0.852 (±0.02)
	Gradient boosting	0.872 (±0.02)	0.704 (±0.14)	0.685 (±0.10)	0.685 (±0.10)	0.872 (±0.02)
	Logistic regression	0.880 (±0.02)	**0.716** (±0.11)	**0.768** (±0.16)	**0.724** (±0.12)	0.880 (±0.02)

Values shown are average (standard deviation).
†In multiclass settings, the precision_μ, recall_μ and F-score_μ are identical to accuracy.

label (majority class with no incident) in the training set. The baseline classifier scored 0.541 (accuracy), 0.541 (recall), 0.292 (precision) and 0.379 (F-score) using an average weighted by support (the number of true instances of the label).

Table 7: Results on the test data set

Features	Model	Acc	R_M	P_M	$F1_M$	$F1_\mu$ [†]
word unigram (1,1)	Support vector machine	0.895	0.689	0.754	0.712	0.895
	Random forest	0.873	0.578	**0.889**	0.647	0.873
	Gradient boosting	0.892	**0.855**	0.734	**0.769**	0.892
	Logistic regression	**0.899**	0.767	0.772	**0.769**	**0.899**
word bigram (2,2)	Support vector machine	0.842	0.539	0.743	0.598	0.842
	Random forest	0.813	0.470	**0.864**	0.534	0.813
	Gradient boosting	0.811	0.547	0.669	0.592	0.811
	Logistic regression	**0.845**	**0.557**	0.761	**0.617**	**0.845**
word unigram (1,1) and word bigram (1,2)	Support vector machine	**0.901**	0.705	0.768	0.730	**0.901**
	Random forest	0.864	0.537	**0.905**	0.612	0.864
	Gradient boosting	0.887	0.746	0.711	0.723	0.887
	Logistic regression	0.899	**0.773**	0.808	**0.789**	0.899

[†]In multiclass settings, the $precision_\mu$, $recall_\mu$ and $F\text{-score}_\mu$ are identical to accuracy.

On the test set, the unigram and bigram logistic regression models outperformed all the unigram or bigram models in terms of $F\text{-score}_M$ (0.789). In terms of accuracy, the unigram and bigram support vector classification model scored the highest (0.901) while the unigram and bigram random forest model scored the highest $precision_M$ (0.905). The unigram gradient boosting model scored the highest $recall_M$ (0.855).

Experiment 2

In the first experiment, we used word n-gram TF-IDF features. The word unigram model obtained the highest training F-score of 0.736 (s.d.=0.11) for the logistic regression classifier. The same classifier also obtained the highest test F-score of 0.789 using word unigrams and bigrams. For the second experiment, we extracted Word2Vec features and combined the word vector model (100 dimension) with the TF-IDF feature space (unigram) of Experiment 1. Again, model selection was performed on the training data set using 10-fold cross-validation in an exhaustive grid search over the same hyperparameter configurations (see Table 6). The parameters for SVM and random forest stayed the same. For gradient boosting, the learning_rate changed from 0.25 to 0.5. For logistic regression, the tolerance for stopping criteria changed from 10 to 1. We report the scores for the event class as we did for Experiment 1 (Table 8).

The logistic regression classifier using unigrams with Word2vec features obtained the highest training $F\text{-score}_M$ of 0.723 (s.d.=0.12). This value is slightly lower than the training $F\text{-score}_M$ of 0.736 (s.d.=0.11) obtained using the word unigram model in Experiment 1. The accuracy and $recall_M$ scores were also marginally lower for the logistic regression classifier, but surprisingly the $precision_M$ score was marginally higher with 0.780 (s.d.=0.16) compared to 0.775 (s.d.=0.15). These results would suggest that adding additional features created by the CBOW and skip-gram Word2vec models did not improve the training performance of the classifiers. We also tested the performance of the four models on the test set, the results of which can be found in Table 9. The models were first tested using embeddings and thereafter embeddings with the unigram model. Again, we included a majority baseline and used the frequency of the majority class (no incident) as accuracy in the training set. The baseline classifier scored 0.541 (accuracy), 0.541 (recall), 0.292 (precision) and 0.379 (F-score) using an average weighted by support (the number of true instances of the label).

For the test set, the logistic regression classifier using unigrams with Word2vec features obtained the highest test $F\text{-score}_M$ of 0.762. This score was marginally lower than the $F\text{-score}_M$ of 0.769 obtained with the same classifier using only a unigram word model in Experiment 1. Additionally, the test $F\text{-scores}_M$ of random forest and gradient boosting classifiers were also lower. Surprisingly, the test accuracy, $recall_M$, $precision_M$ and $F\text{-score}_M$ of the SVM classifier were higher for the unigram word model with Word2vec features. This would suggest that using additional features created by the CBOW and skip-gram Word2Vec models could improve the performance of classification (or prediction) of new data. Figure 2 depicts the accuracy scores of the training set for the algorithms using unigram model with Word2vec features.

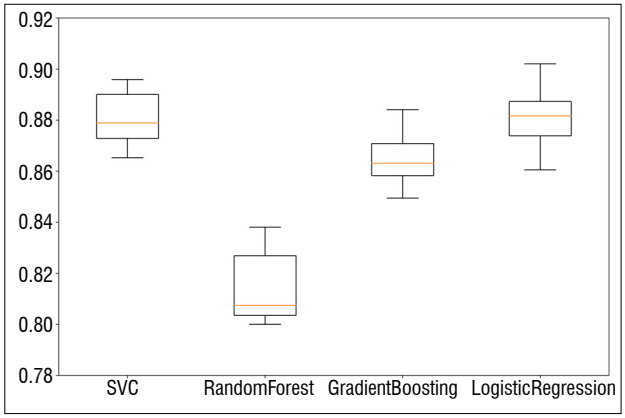


Figure 2: Accuracy scores.

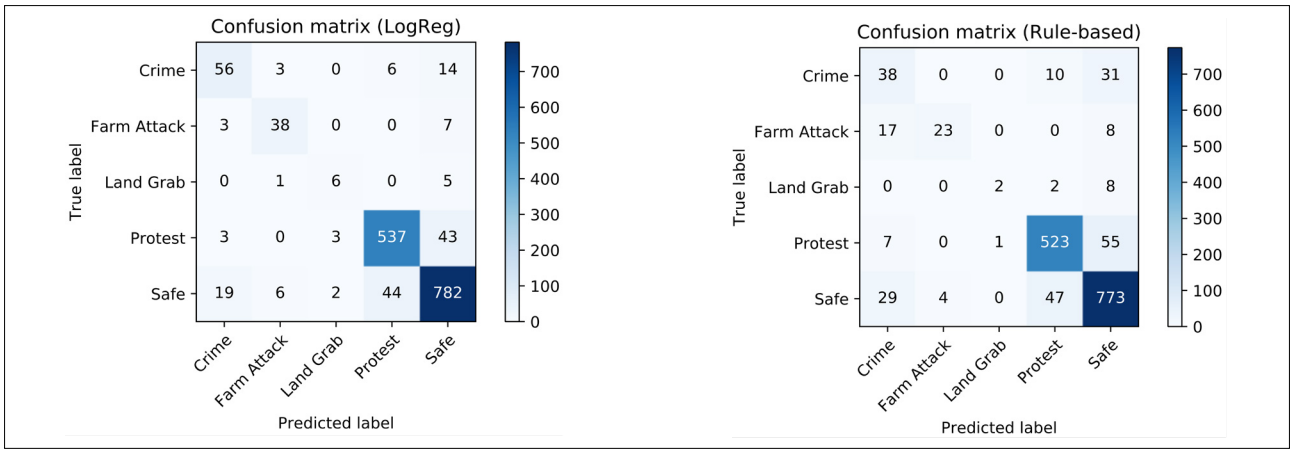


Figure 3: Confusion matrixes.

Table 9: Results on the test data sets

	Features	Model	Acc	R_M	P_M	$F1_M$	$F1_\mu$ †
Experiment 1	word unigram (1,1)	Support vector machine	0.895	0.689	0.754	0.712	0.895
		Random forest	0.873	0.578	**0.889**	0.647	0.873
		Gradient boosting	0.892	**0.855**	0.734	**0.769**	0.892
		Logistic regression	**0.899**	0.767	0.772	**0.769**	**0.899**
Experiment 2	Word2vec (CBOW and skip-gram)	Support vector machine	0.677	0.693	0.460	0.473	0.677
		Random forest	**0.823**	0.498	**0.549**	0.514	**0.823**
		Gradient boosting	0.814	0.530	0.535	**0.530**	0.814
		Logistic regression	0.712	**0.756**	0.485	0.514	0.712
	Word2vec (CBOW and skip-gram) and word unigrams (1,1)	Support vector machine	**0.901**	0.712	0.781	0.739	**0.901**
		Random forest	0.821	0.492	0.790	0.536	0.821
		Gradient boosting	0.873	**0.770**	0.689	0.717	0.873
		Logistic regression	0.897	0.745	**0.787**	**0.762**	0.897

†In multiclass settings, the $precision_\mu$, $recall_\mu$ and $F\text{-}score_\mu$ are identical to accuracy.

Evaluating the machine-learning classifier

In the exploratory phase of this project, we experimented with a rule-based system that classifies messages in the abovementioned four categories based on the occurrence of certain key terms. We compiled dictionaries using unigrams and bigrams and classified messages as 'unsafe' and according to category based on whether these terms appeared (e.g. 'robbery', 'unrest', 'throwing rocks', 'burning tires'). Because no annotated data was available, we could not measure the accuracy of this rule-based classifier. In an effort to better understand the performance of the classifier, we compared our machine-learning classifier with the rule-based system. The machine-learning classifier and the rule-based classifier were both evaluated using the same testing data set (n=1578). The logistic regression classifier (using unigrams) achieved an accuracy of 89.73%, and the rule-based classifier reached 86.12%. The same tendency was observed when comparing the F-scores. The logistic regression classifier obtained an $F\text{-}score_M$ of 0.770 and the rule-based classifier 0.621. The confusion matrixes from the classification test are depicted in Figure 3.

Conclusions and future research

We have presented an approach to automatically classify violent events taking place in South Africa. Our aim was to find the best word model on event detection using WhatsApp messages. Our experiment shows the weakness of these algorithms when applied to a highly imbalanced data set. Additionally, we also used word embeddings created with Word2Vec, using both CBOW (sg=0) and (sg=1) skip-gram models. Our experiments have shown that more work is required to find a universal approach to solving the imbalanced distribution problem in this domain. However, we did train a logistic classifier using unigram, CBOW and skip-gram Word2Vec models that achieved a higher accuracy and F1-score than did the rule-based classifier. Possible future research includes employing Doc2Vec[33], which is an extension of Word2Vec. Pre-trained English and Afrikaans word vectors such as fastText[47] could also be considered, and would thus create a multilingual pre-trained word vector space for the classifiers.

Acknowledgements

We acknowledge the annotation efforts of staff and students at the University of the Free State. We also thank the anonymous reviewers for their helpful comments and advice.

Authors' contributions

E.K.: Project leadership, conceptualisation, methodology, data analysis, data curation, validation, writing – the initial draft, writing – revisions. B.S.: Conceptualisation, data collection, writing – revisions. W.D.: Methodology, validation, writing – revisions.

References

1. Stieglitz S, Mirbabaie M, Ross B, Neuberger C. Social media analytics – Challenges in topic discovery, data collection, and data preparation. Int J Inf Manage. 2018;39(October):156–168. https://doi.org/10.1016/j.ijinfomgt.2017.12.002
2. Gundecha P, Liu H. Mining social media: A brief introduction. In: Tutorals in operations research. Catonsville, MD: INFORMS; 2012. p. 1–17. https://doi.org/10.1287/educ.1120.0105
3. Sarkar D. Text analytics with Python : A practical real-world approach to gaining actionable insights from your data. New York: Apress; 2016. http://doi.org/10.1007/978-1-4842-2388-8
4. Aggarwal CC, Zhai C. A survey of text classification algorithms. In: Aggarwal CC, Zhai C, editors. Mining text data. Boston, MA: Springer; 2012. p. 163–222. https://doi.org/10.1007/978-1-4614-3223-4_6
5. Liu B, Zhang L. A survey of opinion mining and sentiment analysis. In: Aggarwal CC, Zhai C, editors. Mining text data. Boston, MA: Springer; 2012. p. 415–463. https://doi.org/10.1007/978-1-4614-3223-4_13
6. Liu B. Sentiment analysis and opinion mining. Synth Lect Hum Lang Technol. 2012;5(1):1–167. https://doi.org/10.2200/S00416ED1V01Y201204HLT016
7. Sebastiani F. Machine learning in automated text categorization. ACM Comput Surv. 2002;34(1):1–47. https://doi.org/10.1145/505282.505283
8. Ingersoll G, Morton T, Farris A. Taming text: How to find, organize, and manipulate it. Shelter Island, NY: Manning Publications Co; 2013.
9. Sinoara RA, Camacho-Collados J, Rossi RG, Navigli R, Rezende SO. Knowledge-enhanced document embeddings for text classification. Knowledge-Based Syst. 2019;163:955–971. https://doi.org/10.1016/j.knosys.2018.10.026
10. Bueno de Mesquita B. The war trap. New Haven, CT: Yale University Press; 1981.
11. Bueno de Mesquita B, Newman D, Rabushka A. Forecasting political events: The future of Hong Kong. New Haven, CT: Yale University Press; 1985.
12. Bueno de Mesquita B, Stockman FN. European community decision making. New Haven, CT: Yale University Press; 1994.
13. O'Brien SP. Crisis early warning and decision support: Contemporary approaches and thoughts on future research. Int Stud Rev. 2010;12(1):87–104. https://doi.org/10.1111/j.1468-2486.2009.00914.x
14. Abdollahian M, Baranick M, Efird B, Kugler J. Senturion: A predictive political simulation model [document on the Internet]. c2006 [cited 2018 Aug 20]. Available from: http://www.ndu.edu/CTNSP/Def_Tech/DTP 32 S%0Aenturion.pdf
15. Esty DC, Goldstone JA, Gurr TR, Surko P, Unger AN. State failure task force report. McLean, VA: Science Applications International Corporation; 1995.
16. Bates RH, Epstein DL, Goldstone JA, Gurr TR, Harff B, Kahl CH, et al. Political instability task force report: Phase IV findings. McLean, VA: Science Applications International Corporation; 2003.

17. Goldstone JA, Bates RH, Epstein DL, Gurr TR, Lustik MB, Marshall MG, et al. A global model for forecasting political instability. Am J Pol Sci. 2010;54(1):190–208. https://doi.org/10.1111/j.1540-5907.2009.00426.x

18. Raleigh C, Linke A, Hegre H, Karlsen J. Introducing ACLED: An armed conflict location and event dataset. J Peace Res. 2010;47(5):651–660. https://doi.org/10.1177/0022343310378914

19. Sundberg R, Melander E. Introducing the UCDP georeferenced event dataset. J PeaceRes.2013;50(4):523–532.https://doi.org/10.1177/0022343313484347

20. Croicu M, Sundberg R. UCDP GED Codebook version 17.1. Uppsala: Department of Peace and Conflict Research, Uppsala University; 2017.

21. Salehyan I, Hendrix CS, Hamner J, Case C, Linebarger C, Stull E, et al. Social conflict in Africa: A new database. Int Interact. 2012;38(4):503–511. https://doi.org/10.1080/03050629.2012.697426

22. Schrodt PA. Precedents, progress, and prospects in political event data. Int Interact. 2012;38(4):546–569. https://doi.org/10.1080/03050629.2012.697430

23. Gerner DJ, Schrodt PA, Francisco RA, Weddle JL. Machine coding of event data using regional and international sources. Int Stud Q. 1994;38(1):91. https://doi.org/10.2307/2600873

24. Hendrix CS, Salehyan I. No news is good news: Mark and recapture for event data when reporting probabilities are less than one. Int Interact. 2015;41(2):392–406. https://doi.org/10.1080/03050629.2015.982117

25. Anonymous. Study finds Facebook news use declining, WhatsApp growing [Internet]. SABC News. 2018 June 14 [cited 2018 Oct 01]. Available from: http://www.sabcnews.com/sabcnews/study-finds-facebook-news-use-declining-whatsapp-growing/

26. Newman N, Fletcher R, Levy DAL, Nielsen RK. Digital News Report 2018. Oxford: Reuters Institute; 2018.

27. Anonymous. South Africans turn to technology to fight crime [webpage on the Internet]. c2015 [cited 2018 Nov 30]. Available from: http://visiontactical.co.za/2015/07/08/south-africans-turn-to-technology-to-fight-crime/

28. Haddi E, Liu X, Shi Y. The role of text pre-processing in sentiment analysis. Procedia Comput Sci. 2013;17:26–32. https://doi.org/10.1016/j.procs.2013.05.005

29. Dey L, Haque SM. Opinion mining from noisy text data. Int J Doc Anal Recognit. 2009;12(3):205–226. https://doi.org/10.1007/s10032-009-0090-z

30. Bird S, Klein E, Loper E. Natural language processing with Python. Vol. 43. Sebastopol, CA: O'Reilly; 2009.

31. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas. 1960;20(1):37–46. https://doi.org/10.1177/001316446002000104

32. Indurkhya N, Damerau FJ. Handbook of natural language processing. 2nd ed. Boca Raton, FL: CRC Press; 2010.

33. Le QV., Mikolov T. Distributed representations of sentences and documents. In: Proceedings of the 31st International Conference on Machine Learning; 2014 June 21–26; Beijing, China. Red Hook, NY: Curran Associates, Inc.; 2014. p. 1188–1196.

34. Goodfellow I, Bengio Y, Courville A. Deep learning: Machine learning book [homepage on the Internet]. c2016 [cited 2019 Aug 19]. Available from: http://www.deeplearningbook.org/

35. Ram A, Nagappan M. Supervised sentiment classification with CNNs for diverse SE datasets [article on the Internet]. c2018 [cited 2019 Jan 17]. Available from: https://arxiv.org/abs/1812.09653?context=cs

36. Mikolov T, Sutskeve I, Chen K, Corrado G, Dean J. Distributed representations of words and phrases and their compositionality. In: Burges CJC, Bottou L, Welling M, editors. Advances in neural information processing systems 26: 27th Annual Conference on Neural Information Processing Systems; 2013 December 05–08; Lake Tahoe, Nevada, USA. Red Hook, NY: Curran Associates, Inc.; 2013. p. 3111–3119.

37. Mäntylä MV, Graziotin D, Kuutila M. The evolution of sentiment analysis – A review of research topics, venues, and top cited papers [article on the Internet]. 2016;abs/1612.0. Available from: http://arxiv.org/abs/1612.01556

38. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. J Mach Learn Res. 2011;12:2825–2830.

39. He H, Garcia EA. Learning from imbalanced data. IEEE Trans Knowl Data Eng. 2009;21(9):1263–1284. https://doi.org/10.1109/TKDE.2008.239

40. Lemaître G, Nogueira F, Aridas CK. Imbalanced-learn: A Python toolbox to tackle the curse of imbalanced datasets in machine learning. J Mach Learn Res. 2017;18(17):1–5.

41. Raschka S, Mirjalili V. Python machine learning: Machine learning and deep learning with Python, scikit-learn, and TensorFlow. 2nd ed. Birmingham, UK: Packt Publishing; 2017.

42. Chang C, Lin C. LIBSVM. ACM Trans Intell Syst Technol. 2011;2(3):1–27. https://doi.org/10.1145/1961189.1961199

43. Chollet F. Deep learning with Python. Shelter Island, NY: Manning Publications Co.; 2017.

44. Oza NC, Tumer K. Classifier ensembles: Select real-world applications [article on the Internet]. Inf Fusion. 2008;9(1):4–20. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1566253507000620

45. Sokolova M, Lapalme G. A systematic analysis of performance measures for classification tasks. Inf Process Manag. 2009;45(4):427–437. https://doi.org/10.1016/j.ipm.2009.03.002

46. scikit-learn developers. Model evaluation: Quantifying the quality of predictions. No date [cited 2019 Oct 20]. Available from: https://scikit-learn.org/stable/modules/model_evaluation.html

47. Bojanowski P, Grave E, Joulin A, Mikolov T. Enriching word vectors with subword information. Trans Assoc Comput Linguist. 2017;5:135–146. https://doi.org/10.1162/tacl_a_00051

AUTHORS:
Kylie de Jager[1]
Chipo Chimhundu[1]
Yolande X.R. Harley[2]
Tania S. Douglas[1]

AFFILIATIONS:
[1]Division of Biomedical Engineering, University of Cape Town, Cape Town, South Africa
[2]Faculty Research Office, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa

CORRESPONDENCE TO:
Tania Douglas

EMAIL:
tania.douglas@uct.ac.za

DATES:
Received: 12 Mar. 2019
Revised: 30 Nov. 2019
Accepted: 07 Jan. 2020
Published: 26 Mar. 2020

HOW TO CITE:
De Jager K, Chimhundu C, Harley YXR, Douglas TS. Collaboration and citation impact: Trends in health sciences research at the University of Cape Town. S Afr J Sci. 2020;116(3/4), Art. #6132, 8 pages. https://doi.org/10.17159/sajs.2020/6132

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:
bibliometrics, relative citation rate, impact factor, co-authorship

FUNDING:
None

Collaboration and citation impact: Trends in health sciences research at the University of Cape Town

Against a background of substantial growth in publication output in health sciences at the University of Cape Town (UCT) over the past two decades, we examined the relationship between collaboration with domestic and foreign institutions and resulting citations of co-published work. We report on trends in authorship and citation impact for health sciences research at UCT across three 3-year periods: 1999–2001, 2006–2008 and 2013–2015. We examined numbers of collaborative publications with domestic and foreign co-authors; the status of collaboration with other African countries; the location of the 'drivers' of the research (with 'drivers' indicated by first or last authorship); and expected and observed citation counts – used as an indicator of impact – over time. We found that the relative citation rate of the set of UCT health sciences publications has increased; the set of 1999–2001 publications was less frequently cited than expected for the journals in which the publications appear, while the 2006–2008 and 2013–2015 sets were cited more frequently than expected. Relative citation rates were greater for papers for which UCT shared international co-authorship than for papers with UCT-only or domestic co-authorship. Our findings confirm reports in the literature of higher citation of internationally co-authored publications. We additionally found that the publications with the highest relative citation rates were driven by authors from foreign institutions.

Significance:

- Methods are presented for extracting, measuring, analysing and representing the citation impact of collaborative research.
- The relative citation rate of health sciences publications produced by UCT has increased and co-publication with international authors has increased.
- The findings confirm reports in the literature of higher citation of publications co-authored with international collaborators.
- An apparent influence of foreign drivers on citation impact, holds risk for South African science.

Introduction

International collaboration for health-related research is encouraged by governments, funding agencies and university executives, and is sought by researchers. Motives for collaboration include access to equipment, infrastructure, knowledge, expertise and funding, as well as raising research and researcher profiles. In addition, bibliometric research showing associations between international co-authorship and research quality indicators[1,2] creates the expectation that international partnerships increase research impact.

Since the first democratic elections in 1994, the South African government has emphasised the development of science and innovation policy, in an effort to use science and technology as levers for socio-economic development.[3] South African funding agencies extol the benefits of partnership. The South African Medical Research Council, the National Research Foundation, and the Department of Science and Technology, in their recent strategic plans, encourage collaboration and international partnership.[4-6] These agencies also co-fund international research partnerships with foreign agencies such as the US National Institutes of Health and the UK Medical Research Council. A study comparing South Africa's publication volumes for the periods 1990–1994 and 2004–2008 has suggested that the substantial increases in the latter period could mainly be attributed to the strengthening of collaboration with foreign partners.[3] More recently, in a report on the state of research in South Africa, Mouton et al.[7] have shown that, among papers published by South African authors, the proportion of papers co-published with collaborators in Africa and in the rest of the world has increased steadily over the period 2000–2016, while single-author papers and those with South African collaborators only, have declined. Mouton et al.[7] also report that South Africa's publication output since 2000 has shown an average growth rate of 2.9% annually, with the country's contribution to global research output increasing from 0.4% in 2000 to 0.91% in 2016. The Mouton et al.[7] report further shows that the citation impact of South African publications has increased over the period covered by their analysis. Citations are a measure of the acknowledgement by researchers of the work published by their peers.

At South Africa's University of Cape Town (UCT), the annual number of publications in health sciences (journal articles and reviews) listed in Scopus[8] has increased from 408 to 1729 between 1999 and 2015. Little is documented about the nature of the international collaborations that have contributed to these publications, and whether and how they have changed over time. The aim of our study was to examine the relationship between these collaboration patterns and the citation impact of the health sciences research outputs of UCT. We report on trends in collaboration, authorship and citation impact for health sciences research at UCT across three 3-year periods: 1999–2001, 2006–2008 and 2013–2015, using co-publication as an indicator of collaboration.[1] We examined numbers of collaborative publications with domestic and foreign co-authors over the periods studied, the status of collaboration with other African countries, the location of the 'drivers' of the research, and expected and observed citation counts – used as an indicator of impact – over time, from the perspective of a research-active university with considerable access to international collaboration.

Methods

We were interested in publications generated by UCT in the broad field of health sciences. Relevant publications were sourced using Scopus. Titles in Scopus are classified under four broad subject clusters: health sciences; physical sciences; life sciences; and social sciences including humanities. These clusters are further divided into 27 major, and more than 300 minor, subject areas.[9] Of the four broad subject clusters, the health and life sciences clusters, defined by Scopus to consist of 11 major subject areas, contain the publications of interest.

Articles in the health sciences and life sciences clusters with UCT-affiliated authors were extracted by searching Scopus for versions of the name "University of Cape Town", limiting the search to articles and reviews as document types and only considering publications from three time periods: 1999–2001, 2006–2008 and 2013–2015. The search string for 1999 was: AFFILORG("U* of Cape Town" OR "U* of capetown" OR "U* Cape Town" OR "U* capetown" OR uct) AND DOCTYPE (ar OR re) AND PUBYEAR = 1999 AND SUBJAREA (mult OR medi OR nurs OR vete OR dent OR heal OR agri OR bioc OR immu OR neur OR phar). A Scopus search was carried out for each year of the three time periods.

The publication lists identified by Scopus were exported along with citation data (C – all citations; CX – excluding self-citations) and PubMed ID (if present) for each publication. The search was carried out on 18 August 2017 for all publication years of interest. The data for 2014 showed <1% difference between C and CX data (expected to be ~20%). A new Scopus search was conducted on 15 May 2018 for 2014 only and corresponding publication and citation data exported.

Despite limiting our article search to two of the Scopus clusters, the extraction yielded a topic coverage wider than our area of interest. Therefore, to focus the data set, we retained from our Scopus search only those publications that had a PubMed ID. PubMed's subject coverage includes our area of interest only – biomedicine and health – as recorded in the MEDLINE database. Unlike Scopus, PubMed does not provide citation data, hence the use of both databases to extract all relevant data (including citation data), while limiting the data to our area of interest using the PubMed ID.

Our analysis also required SNIP (source normalised impact per publication) and IPP (impact per publication) journal indicators for the three time periods; this information was obtained from the CWTS Journal Indicators website.[10] These indicators have been calculated based on the Scopus bibliographic database.

Data processing

The publication and citation data were exported by Scopus in separate files, as were the CWTS Journal Indicators data. For each publication, Scopus provides a list of author names only, a list of author names with their respective affiliation data, affiliation data only reported as a unique list, a 'corresponding author' name and affiliation list, article title, PubMed ID (PMID), Scopus electronic ID (EID), journal title, International Standard Serial Number (ISSN), volume, issue and page numbers. The CWTS data provide journal title, ISSN, volume and issue number in addition to the SNIP and IPP values.

MATLAB (Mathworks, Natick, MA, USA) was used to perform string comparisons (publication year, ISSN, journal title, volume, issue, article title) on the exported bibliographic and citation information. Matched publications were merged into a single spreadsheet. Discrepancies were flagged, manually checked and corrected. It was not always possible to match each publication with the CWTS Journal Indicators; such publications were excluded from further analysis.

MATLAB was also used to check selected information for each publication, and any inconsistencies were flagged for manual inspection and cleaned:

1. Duplicate article titles, PMIDs and EIDs were identified.
   - Multiple publications with identical PMIDs were found to have identical titles, digital object identifiers (DOIs) and journals (source title, volume, issue and page numbers). Citation counts, however, could be different. Such publications were sourced in Scopus using their title. The citation list for each publication was extracted and compared. Typically, the publications did not have overlapping citations. Consequently, the two duplicate entries were merged into a single entry (title, journal, DOI, PMID) and the total citation count retained (i.e. citation counts for the duplicate publications were added).
   - Duplicate titles were kept as separate entries if they had different DOIs, journals (i.e. differences in source, volume, issue or page number) and PMIDs. Citation counts were retained as separate counts for such duplicates.
   - No duplicate EIDs were found.
2. The list of affiliations associated with each publication retrieved from Scopus was compared with the list of affiliations provided for each co-author. Both lists were then independently examined for occurrences of known country names, and the identified countries from both lists compared. Any discrepancies found through the comparison, as well as any publications that did not contain 'South Africa' in the affiliation lists, were flagged for manual inspection.
   - Online databases (PubMed, Scopus, Google) were used to check affiliation data for the relevant publications as well as typographical errors or missing country information, for manual correction.
   - Some publications were found to not have a South African affiliation (typically due to an incorrectly identified UCT affiliation, see Point 3 below). Such publications were removed from the data set.
3. Variations of the name University of Cape Town (as used in the Scopus search terms) were used to flag publications that contained affiliation names which did not definitively represent the University of Cape Town. For instance:
   - UCT – could be an abbreviation for something other than the University of Cape Town.
   - U of Cape Town – the U could stand for something other than University
   - Unit Cape Town – does not match "University of Cape Town"

   Such publications were manually checked. Non-"University of Cape Town" publications were removed from the data set.
4. The three Scopus lists – author names only, authors with affiliation data and affiliation data only – were compared and differences were flagged for manual inspection to correct for inconsistencies:
   - Publications were found which did not include a separate affiliation for each author. In such cases, the original publication was consulted, and the affiliation list corrected accordingly.
   - In instances in which Scopus affiliation data were missing for a middle author, the original publication was also consulted and the affiliation list corrected accordingly. If the original publication was still found to not show affiliation information for the author, any one of the affiliations of other authors would be used instead for completeness, as it would not affect the driver analysis or country representation information as provided by the publication.
   - If the only affiliation provided was that of the corresponding author for both Scopus data and the original publication, it was assumed that the affiliation was the same for all authors.

Co-authorship trends

Co-author countries were extracted from the affiliation data from each publication. The locations of authors for UCT's health sciences publications were indicated on a geographic heat map using Tableau Public, a freely available data visualisation software tool.[11]

Research drivers

Authorship order usually indicates the level of contribution for each author listed on a publication. In the health sciences and public health fields, the first author is typically the one making the largest contribution, with the last author usually having a more senior or supervisory role, often contributing to the inception of the research project.[12] Lead authorship is determined by the level of responsibility for the manuscript, research contribution and in some cases the responsibility of correspondence after publication. Although there is no universal rule for author listing of publications, often the first and the last authors have a leading role in the direction of the work. As such, this study considered an author to be a driver or leader of the research if they appeared as the first or last author.

In some cases, authors may simply be listed alphabetically without considering level of contribution. To determine the proportion of such cases, publications with alphabetical author lists, where the publication had four or more co-authors, were identified and counted. Due to the small percentage of such publications, the impact of these publications on investigating research driver patterns was considered to be negligible, and the publications were included in the analysis.

The countries driving the research were considered those with which the first and last authors are affiliated; in the case of authors with multiple affiliations from different countries, all such countries were considered to be drivers. The association between driver location and citations was examined.

Publication sets

For each of the three time periods considered (1999–2001, 2006–2008, 2013–2015), the data were separated into mutually exclusive affiliation and driver sets as outlined in Table 1. The affiliation sets considered the affiliations of all co-authors present in a publication, while the driver sets (subsets of the affiliation sets) only considered the affiliations of the first and last authors. The affiliation sets described in Table 1 were defined to determine whether co-authorship included UCT authors only, or domestic (within South Africa) or international partnerships as defined by the U, D and I affiliation sets, respectively. All combinations of driver sets were created as a means of identifying which of the co-authorship combinations within each affiliation set had more influence on the research conducted. Dominance of the IUd publications in the international set, for example, would communicate that, although there is international collaboration, UCT tends to lead the collaborative activity, while a dominance of IFd would suggest that international publications with UCT are largely driven by foreign entities.

Citation impact

Questions concerning whether self-citations should be excluded from citation analyses have been raised[13,14], as self-citations are known to have an impact on certain types of analyses. Definitions of self-citation vary slightly depending on the nature of the data, but, in most cases, self-citation is considered an instance in which both the citing and the cited paper have at least one author in common.[13] This situation is more likely to occur with highly collaborative publications due to the higher number of co-authors. However, the more authors present on a publication, the more likely it is to be cited in general, at two additional citations on average per additional author, with the increasing number of self-citations making a small contribution to increased citation rates.[13]

Some scholars agree that self-citations are problematic for low-level analysis such as at the level of the institution[13,14], and that self-citations have less impact for analyses involving larger groups such as the comparison of citation impact across countries. The share of self-citations in a data set influences whether the exclusion of self-citations should be considered prior to further analysis. Self-citation rates vary across fields due to differences in citation norms and tend to be low in fields such as clinical medicine.[13] Country-level analyses have been conducted using data sets with self-citation shares between 24% and 28%.[2] However, proportions of up to 20% are generally considered permissible for the inclusion of self-citations in citation impact assessments.[13]

The share of self-citations across our studied time periods was between 19% and 22%, which is comparable to previous work that has used data with self-citations ranging from 20% to 28% for country-level analyses.[13,14] Our analysis is at the international level, for a broad scientific field, so is unlikely to be affected significantly by the inclusion of self-citations. However, we chose to exclude self-citations to enable additional comparison of the smaller UCT and domestic publication sets to the much larger international publication sets.

For a publication set, the mean observed citation rate (MOCR) is the total number of citations accumulated in Year Y for publications of Years Y-1, Y-2 and Y-3, divided by the total number of publications, *n*, in that same 3-year publication window. Mean expected citation rate (MECR) is the average number of citations per paper accumulated in Year Y for the journals represented in the publication set, for the same 3-year publication window.[2,15]

The MECR and MOCR, both excluding self-citations, were calculated for all publication sets considered, and were used to compare citation impact across the affiliation and driver publication sets. MOCR and MECR (Equations 1 and 2) were modified for our data set from the

Table 1: Definition of affiliation and driver publication sets

Affiliation set		Description	Author affiliation		Driver set	
			First	Last		
U	UCT	All co-authors have only a UCT affiliation	U	U	U	UCT driven
D	Domestic	Co-authors are affiliated with UCT and at least one South African (non-UCT) affiliation. A single co-author can have multiple affiliations to both UCT and the other South African entity.	U	U	DUd	Domestic-UCT driven
			D	D	DNd	Domestic-non-UCT driven
			U and D in combination		DCd	Domestic-UCT co-driven
I	International	Co-authors are affiliated with UCT and at least one foreign (non-South African) entity. A single co-author can have multiple affiliations including an affiliation with another South African (non-UCT) entity.	U	U	IUd	International-UCT driven
			D	D	IDd	International-domestic driven
			F	F	IFd	International-foreign driven
			U and D in combination		IUDd	International-UCT-domestic co-driven
			U and F in combination		IUFd	International-UCT-foreign co-driven
			D and F in combination		IDFd	International-domestic-foreign co-driven
			U, D and F in combination		IMd	International multi-driven

U, UCT affiliation; D, non-UCT South African (domestic) affiliation; F, non-South African (foreign) affiliation

definitions in previous studies[2,15] through the use of a 3-year publication window instead of a 2-year window, and the use of the Scopus IPP instead of the Clarivate Analytics Journal Citation Reports (JCR) Garfield factor (also known as the journal impact factor).

$$MOCR = \frac{\sum_{i=1}^{n} CX_i}{n} \quad \text{Equation 1}$$

$$MECR = \frac{\sum_{i=1}^{n} IPP_i}{n} \quad \text{Equation 2}$$

IPP differs from the Garfield factor mainly in that IPP is calculated on the basis of papers published in the previous 3 years (Y-1, Y-2 and Y-3) instead of the previous 2 years. The IPP of Year Y, based on a 3-year publication window, would therefore mean collecting citations in Year Y for papers published in Years Y-1, Y-2 and Y-3. As an example, citations for papers published in 2006, 2007 and 2008 would be collected to evaluate the IPP for the year 2009. The longer publication window used by Scopus compared to the window used by Journal Citation Reports is believed to help reduce the impact of differences in citation trends between fields and/or journals.[16]

The MOCR/MECR ratio, known as the relative citation rate (RCR), is a more precise measure of impact than journal-based metrics, as it accounts for citation trends of various fields. It is field-independent and indicates if a publication has been cited as expected.[17] However, the ratio retains an element of journal-level influence; this influence could be eliminated with normalisation by all publications in the field, for which data were not readily available, rather than expected journal citation rate. While the RCR is not explicitly used in this study, citation impact was investigated by plotting MOCR against MECR in a relational chart, as suggested by Schubert and Braun[18]. The MOCR=MECR (or RCR =1) line is the line at which the observed citation rate for the publication set meets the expected citation rate for the journals in which the articles of the publication set appear. The line can be regarded as the boundary between lower and higher than expected impact, but 'not without an element of arbitrariness'[18].

Results

Table 2 shows the attributes of the UCT (U), domestic (D) and international (I) affiliation publication sets. An upward trend in the number of publications in each set is evident, with an approximate doubling of outputs from one 3-year window to the next for the retained data set. For the U, D and I data sets, the number of publications is seen to increase by factors of 1.6, 5.8 and 8.0, respectively, between 1999 and 2015. While an increase in the number of UCT authors is consistent with the overall increase in number of authors per publication, the overall proportion of UCT authors per publication as well as the average number of UCT authors per publication have generally decreased with time. This finding is also in agreement with the downward trend in the UCT affiliation set, which constitutes 42% of the retained publications for 1999–2001, 20% for 2006–2008, and only 13% for 2013–2015.

While very few papers were published in collaboration with other African countries during the 1999–2001 period (eight countries), a general upward trend in African collaboration is observed, with a marked increase between the first two periods such that the 2006–2008 period had four times the number of African collaborators than did the 1999–2001 period.

Co-authorship trends

Figure 1 shows the locations of countries that have co-authored health sciences publications with UCT in the 1999–2001, 2006–2008 and 2013–2015 periods. Each country is colour coded, with cool colours representing few co-publications and hot colours representing a high number of co-publications with UCT. Figure 2 is similar but focuses on the African continent, specifically to show African co-authorship trends. Figures 1 and 2 therefore show two heat maps, representing global collaborations (green to red scale) and African collaboration (light blue to dark blue scale).

In both Figure 1 and Figure 2, an increase in co-publication is observed, consistent with the values in Table 2. The global maps (Figure 1) suggest that co-publication has occurred most frequently with North America, Europe and Australia, and has also increased over time, while the African maps (Figure 2) suggest that intensity of co-publication has increased, particularly with countries in southern and East Africa, with Cameroon, Nigeria and Ghana also showing increased co-publication with UCT over time. As time passed, the African countries present in earlier periods persisted in collaborating with UCT; the increasing co-publication rates may also be the result of new collaborations having been fostered.

Table 3 shows the top 10 countries worldwide (excluding South Africa) that have co-published with UCT, ranked by number of health sciences co-publications. The top two positions have been maintained by the USA and the UK. Most of the countries in the 3rd to 9th places have persisted but with a shuffling of positions.

Table 2: Overview of UCT health sciences publications for the 1999–2001, 2006–2008 and 2013–2015 periods

	Publication set attributes	1999–2001	2006–2008	2013–2015
Publications	Total number of publications	1184	2535	4844
	Proportion of self-citations in retained publications*	19%	22%	22%
	Total number of retained publications (with IPP, SNIP and PMID values) (% of total number of publications)	656 (55%)	1684 (66%)	3298 (68%)
	Total citations of retained publication set, excluding self-citations	1940	8392	17 413
	UCT publications (% of retained publications)	273 (42%)	345 (20%)	442 (13%)
	Domestic publications (% of retained publications)	96 (15%)	320 (19%)	558 (17%)
	International publications (% of retained publications)	287 (44%)	1019 (61%)	2298 (70%)
Authors	Number of authors	3166	9829	30 750
	Average number of authors per publication	5	6	9
	Number of UCT authors	1838	4149	8007
	Average number of UCT authors per publication	3	2	2
	Proportion of UCT authors	58%	42%	26%
	Proportion of publications with four or more authors in alphabetical order	1.1%	1.0%	0.5%
Countries	Number of countries	51	99	148
	Number of African countries (excluding South Africa)	8	32	43

IPP, impact per publication; SNIP, source normalised impact per publication; PMID, PubMed reference number
**Retained publications had complete affiliation data, PubMed IDs, CWTS Journal Indicators data and were also verified to have at least one UCT author amongst the listed authors.*

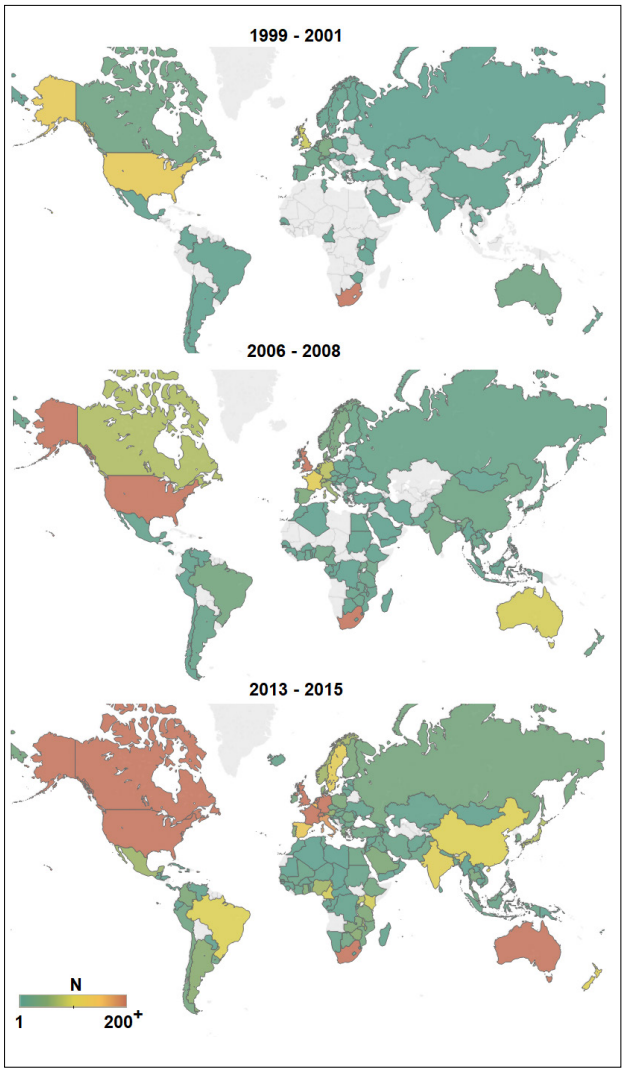


Figure 1: Global footprint of UCT health sciences co-authors. Number of co-publications (N) increase in the direction green to red, with co-publications of 200 or more shown in red.

Table 3 also shows the number of co-publications with countries on the African continent (excluding South Africa) – as these were very few in the first time period, the ranking display is limited to four ranks. For 1999–2001, the highest number of co-publications observed in the data set was two. As African collaboration increased with time, ranks became more clearly defined with the top four ranking countries being Nigeria, Uganda, Kenya and Cameroon in the latter two time periods, although positions varied.

Figure 3 shows the numbers and proportions of publications in each of the publication sets defined in Table 1, with the U, D and I sub-totals matching those listed in Table 2.

In accordance with Table 2, Figure 3 shows that the proportion of UCT-only authored papers has decreased over time. This decrease is accompanied by a substantial increase in internationally co-authored papers. With respect to the drivers of the research (first/last authorship), in the domestic sub-category, the domestic co-driven (DCd) publications are more prevalent than the UCT-only driven (DUd) and the domestic non-UCT driven (DNd) ones. At the international level, the foreign-only driven papers (IFd) and those co-driven by UCT and a foreign entity (IUFd), constitute the largest contribution to the international co-publications, together accounting for (in roughly equal measures) approximately 72% of the international publication set over the studied period.

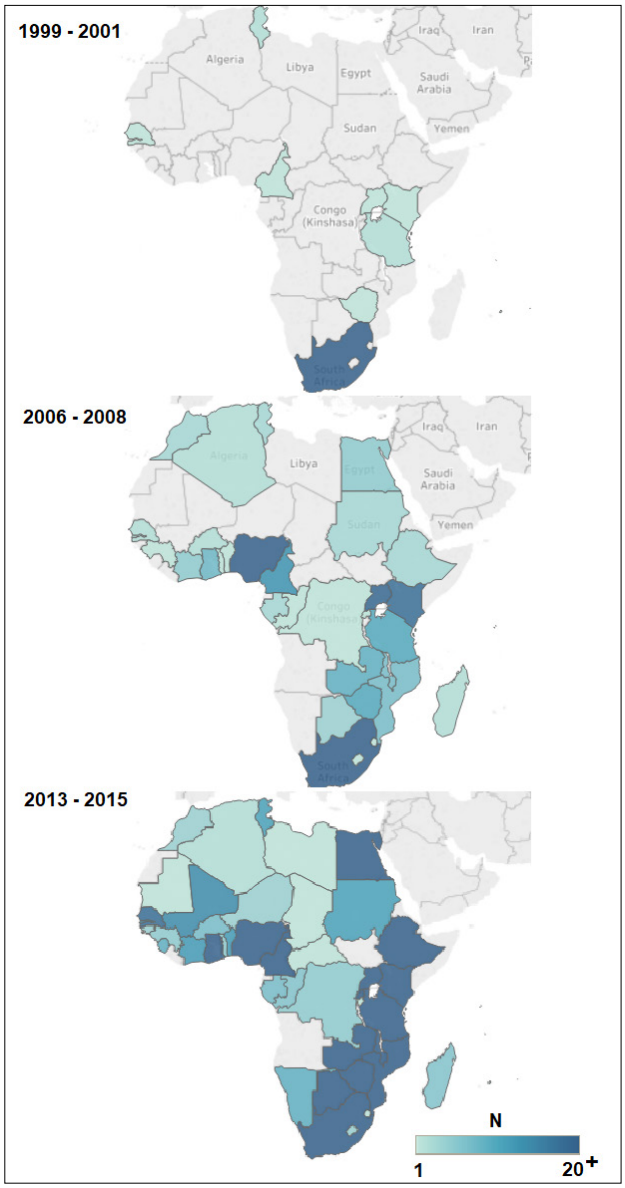


Figure 2: African footprint of UCT health sciences co-authors. Number of co-publications (N) increase in the direction light blue to dark blue, with co-publications of 20 or more shown in dark blue.

Citation impact

The documents with IPP were used for MOCR and MECR calculation. Figure 4 plots MOCR against MECR, with each marker representing a publication set. The publication sets are shown in relation to the MOCR=MECR line. As time progresses, most of the markers move closer to the line, except for the IFd subset (differentiated from all other subsets with a diamond marker). The IFd subset is seen to move upward, rightward and further from the MECR=MOCR line.

Discussion

Co-publication is a collaborative activity[2,19,20], and the qualifier for authorship is all authors contributing, in various degrees, to the design, investigation, manuscript writing and approval of the final product[12]. Co-publication can therefore be seen as a proxy for collaboration, although this relationship can be distorted.[21] Our data set shows a notable increase over time in the number of health sciences publications (Table 2), consistent with the observation that, in general, South Africa's scientific publication output has climbed steeply since around 2004[7] after fairly constant levels had been maintained from the mid-1980s[3]. The rate at which the international publication set increases is higher than that for the UCT and domestic sets.

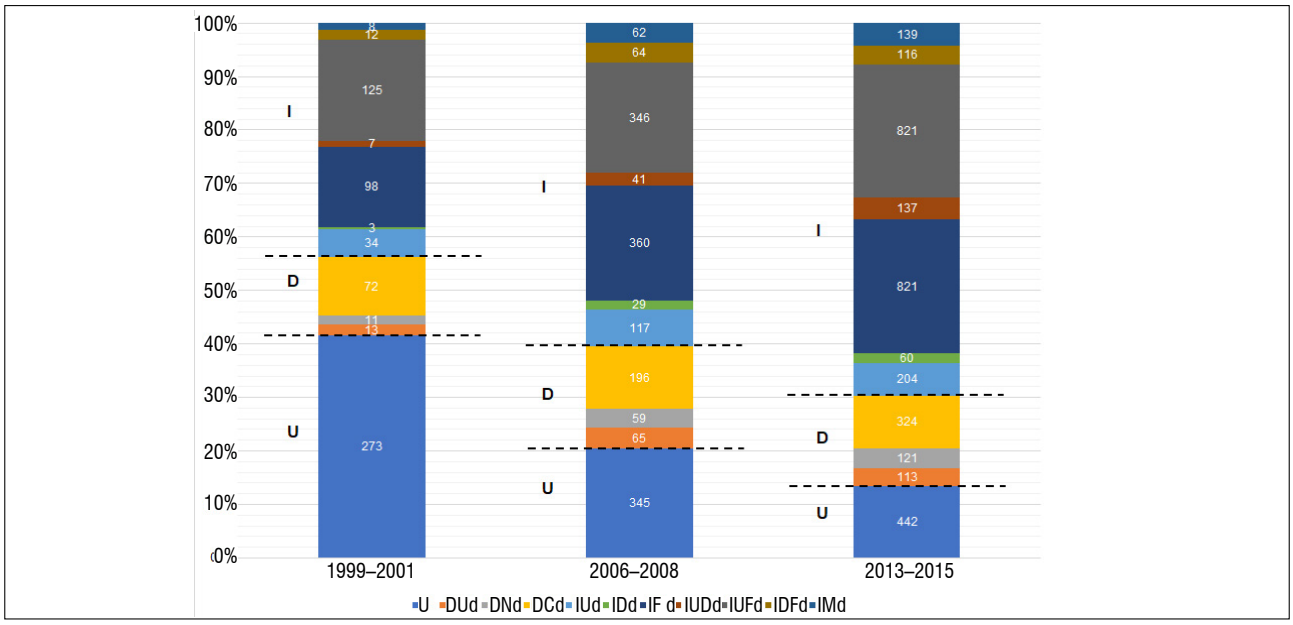
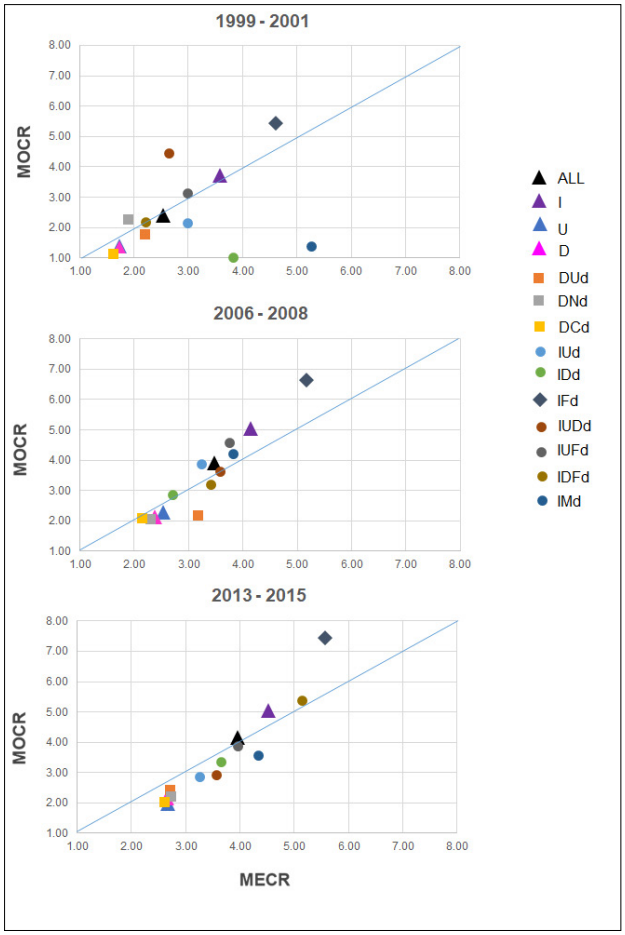


Figure 3: Breakdown of UCT health sciences publication driver sets over time. The number of publications is shown on the bars.



MOCR, mean observed citation rate; MECR, mean expected citation rate
Key as per Table 1

Figure 4: Relational charts over time for UCT health sciences publications.

Coupled with a growth in volume of output, UCT's health sciences research has therefore shifted to increased domestic and especially international co-publication. UCT's collaboration with African countries, specifically, has also grown. Mouton et al.[7] have shown an increase in international co-publication for South African research outputs in health sciences, from 59% in 2000 to 64% in 2016, whereas our data set shows an increase in international co-publication from 44% for 1999–2001, to 70% for 2013–2015 (retained publications in Table 2).

The Mouton report attributes high international collaboration rates for health sciences to internationally supported clinical trials and other large projects. Mouton et al.[7] also describe the existence of 'mega-research' institutes, which are substantially funded by international organisations, as a distinctive feature of the South African health research landscape. These facilities have the potential to expand the country's health research competencies and leadership and the ability of its researchers to drive collaborative research agendas. The assembly of effective and productive local research teams is not, however, limited to such large facilities. Further analysis of the composition and contributions of co-authors of publications with high citation impact, might reveal a range of strategies for building successful collaborative teams.

Table 2 shows that the number of authors of any affiliation in the full set of publications has increased at a faster rate (approximately 1:3:10) than the number of authors with UCT affiliations specifically (approximately 1:2:4). This reflects the increase in external collaborators, with the observed proportion of UCT authors across the publication sets decreasing from 57% to 36% across the studied periods. In addition, while the total number of UCT authors has increased with time, the average number of UCT authors per publication has remained around 2. As also evidenced in Table 2, the number of citations (excluding self-citations) has increased at a faster rate (approximately 1:4:9) than has the number of retained publications (approximately 1:3:5) across the three studied periods (1999–2001:2006–2008:2013–2015). As the number of citations generally increases at a rate of two citations per additional author[13], the faster growth in citations is expected.

UCT's co-authors show a wide geographical spread across continents. There is a general increase in collaboration worldwide, including with other African countries, as shown in Figures 1 and 2. Throughout the studied period, the top two countries collaborating with UCT have been the USA and the UK (Table 3). Sooryamoorthy[22] similarly found the USA and the UK to be the leading co-publishing countries with South Africa for papers in the Web of Science for the period 1945–2010. UCT's top nine co-publishing countries have remained consistent over the 16-year period, comprising mostly European countries.

Where UCT's co-publication is purely domestic, our findings reveal that it is more common for UCT and a local entity to jointly drive the research. However, with international collaboration, the more likely nature of co-publication is for either a foreign entity to completely drive the research

Table 3: Top countries (excluding South Africa) by number of health sciences co-publications with the University of Cape Town

	1999–2001		2006–2008		2013–2015	
Rank	**Country**	**Number of publications**	**Country**	**Number of documents**	**Country**	**Number of publications**
1	USA	124	USA	465	USA	1125
2	United Kingdom	92	United Kingdom	381	United Kingdom	836
3	Germany	34	France	113	Netherlands	296
4	Switzerland	24	Australia	96	Australia	295
5	Australia	23	Netherlands	82	France	261
6	Canada	22	Germany	81	Germany	260
7	Netherlands	22	Canada	76	Switzerland	256
8	France	16	Switzerland	64	Canada	252
9	Italy	15	Italy	63	Italy	179
10	Spain	13	Denmark	41	Belgium	154
Top African countries						
1	Cameroon, Tanzania, Tunisia	2	Nigeria	20	Cameroon, Kenya	95
2			Uganda	19		
3			Kenya	18	Uganda	80
4	Kenya, Mauritius, Senegal, Uganda, Zimbabwe	1	Cameroon	13	Nigeria	68

or for UCT to jointly drive the research with the foreign entity (Figure 3). Foreign research drivers may be supported by funding from non-South African agencies, and these publications would then reflect projects on which UCT researchers are co-investigators rather than principal investigators. The growth in UCT-foreign co-driven publications may reflect an increase in access to international funding by UCT researchers on projects that they co-lead with foreign researchers. Examination of funding attributions on publications would clarify funding sources for different types of co-publication.

The full set of UCT publications ('All' in Figure 4) lies below the MOCR=MECR line in the first time period, but above the line in the second and third periods. This pattern indicates an increase in the relative citation rate of UCT publications from lower than to higher than expected (for the journals in which the publications appear) over time. Identification of the publications with the highest relative citation rates would enable analysis of the topics receiving high citation attention and consideration of the health sciences impact of such publications, as well as deeper examination of the nature of the collaboration involved. Such analysis was beyond our scope but presents an area for future study.

Foreign partnerships strengthen UCT's citation impact: in all three time periods the internationally collaborative publication set (I) is above the MOCR=MECR line, whereas U and D are below the line. Thus, the relative citation rate is greater for papers on which UCT has international co-authorship than it is for papers with UCT-only or domestic co-authorship. The mean observed citation rate is also higher for I than it is for U or D in all time periods, indicating that UCT's internationally collaborative publications typically have a higher citation rate than internal or domestic publications. This finding supports what has been described in the literature. For example, Abramo et al.[1] found a correlation between journal impact factor and co-publication with foreign co-authors in the Italian research system. Mouton et al.[7] found an increase in the mean normalised citation score of South Africa's health sciences papers over the period 2000–2016, from 0.9 to 1.3; this increase means that South Africa's health sciences publications have on average shifted over time from being cited slightly less frequently to more frequently than the world average in the relevant fields. The growth in citation impact for the UCT health sciences example reflects the situation in South Africa more broadly, suggesting that increased citation impact at the national level may at least partly be due to increased international collaboration, consistent with findings in other countries.[1,2]

Higher citation of internationally co-authored publications may be attributed to an audience effect, in which more authors from a greater range of countries provide access to a larger community of citing researchers, and the growth in international collaboration may reflect preferential attachment to international co-authors based on their status and reputation.[23] Kahn[3] expressed concern that South African science might be vulnerable should there be weakening of collaborations with foreign partners, given that foreign collaboration had likely driven increases in publication volumes in recent years. However, in a study using data from Scopus and Web of Science, Wagner et al.[23] found evidence to suggest that international collaborations suppress novelty and produce conventional outputs. The authors cite obstacles generated by collaboration, such as communication barriers and costs that limit creativity, as possible reasons for this finding. If science and technology are to be drivers of innovation and socio-economic development in South Africa, a lack of novelty in actively encouraged and rapidly proliferating international collaborations is a concern.

The motivations for international collaboration proposed by Wagner et al.[23] do not specifically address the driving authors of collaborative research. As indicated by the highest MOCR in all three time periods, UCT publications that are internationally collaborative and driven by the foreign entity (IFd) tend to be cited more frequently than publications with other driver combinations, and are also cited at higher rates than expected for the journals in which these papers are published (Figure 4, diamond marker). Despite the growth in publication productivity, the apparent influence of foreign drivers on, and the likely associated dependence on foreign principal investigators for high citation impact, holds risk for South African science.

Acknowledgements

We thank Dr Robyn May and Ms Esmari Huysamen for early assistance with data collection and analysis.

Authors' contributions

T.S.D. led and managed the project. All authors contributed to conceptualisation. K.d.J. collected the data. K.d.J. and C.C. developed the methodology, curated the data and analysed the data. C.C. validated the data. K.d.J. and C.C. wrote the initial draft. All authors contributed to the writing of the final manuscript.

References

1. Abramo G, D'Angelo CA, Di Costa F. Research collaboration and productivity: Is there correlation? High Educ. 2009;57(2):155–171. https://doi.org/10.1007/s10734-008-9139-z
2. Wang L, Thijs B, Glänzel W. Characteristics of international collaboration in sport sciences publications and its influence on citation impact. Scientometrics. 2015;105(2):843–862. https://doi.org/10.1007/s11192-015-1735-y
3. Kahn MJ. A bibliometric analysis of South Africa's scientific outputs – some trends and implications. S Afr J Sci. 2011;107(1/2), Art. #406, 6 pages. https://doi.org/10.4102/sajs.v107i1/2.406

4. South African National Research Foundation. National Research Foundation strategy 2020 [document on the Internet]. c2019 [cited 2019 Nov 04]. Available from: https://www.nrf.ac.za/sites/default/files/documents/NRF%20Strategy%20Implementation.pdf
5. South African Medical Research Council. SAMRC strategic plan, for the fiscal years 2015/16–2019/20 [document on the Internet]. c2015 [cited 2019 Nov 04]. Available from: http://www.mrc.ac.za/flipbooks/StratPlan/mobile/index.html#p=1
6. South African Department of Science and Technology. Department of Science and Technology strategic plan for the fiscal years 2015–2020 [document on the Internet]. c2015 [cited 2019 Nov 04]. Available from: https://www.dst.gov.za/images/Attachments/DST_Strategic_Plan_2015-2020_web_pdf.pdf
7. Mouton J, Basson, I, Blackenberg J, Boshoff N, Prozesky H, Redelinghuys H, et al. The state of the South African research enterprise. Stellenbosch: DST-NRF Centre of Excellence in Scientometrics and Science, Technology and Innovation Policy; 2019. Available from: http://www0.sun.ac.za/scistip/wp-content/uploads/2019/08/state-of-the-South-African-research-enterprise.pdf
8. Elsevier. Scopus [database on the Internet]. c2004 [cited 2019 Mar 06]. Available from: https://www.scopus.com/
9. Elsevier B.V. Scopus: Content coverage guide [document on the Internet]. c2017 [cited 2019 Nov 04]. Available from: https://www.elsevier.com/__data/assets/pdf_file/0007/69451/0597-Scopus-Content-Coverage-Guide-US-LETTER-v4-HI-singles-no-ticks.pdf
10. Leiden University's Centre for Science and Technology Studies. CWTS Journal Indicators [homepage on the Internet]. c2018 [cited 2019 Mar 05]. Available from: http://www.journalindicators.com/
11. Tableau. Maps [software on the Internet]. c2003 [cited 2019 Nov 04]. Available from: https://www.tableau.com/solutions/maps
12. Tarkang EE, Kweku M, Zotor FB. Publication practices and responsible authorship: A review article. J Public Health Africa. 2017;8(723):36–42. https://doi.org/10.4081/jphia.2017.723
13. Aksnes DW. A macro study of self-citation. Scientometrics. 2003;56(2):235–246. https://doi.org/10.1023/A:1021919228368
14. Glänzel W, Thijs B. The influence of author self-citations on bibliometric macro indicators. Scientometrics. 2006;59(3):281–310. https://doi.org/10.1023/B:SCIE.0000018535.99885.e9
15. Vinkler P. Relations of relative scientometric indicators. Scientometrics. 2003;58(3):687–694. https://doi.org/10.1023/B:SCIE.0000006888.69146.24
16. Moed HF. Measuring contextual citation impact of scientific journals. J Informetr. 2010;4(3):265–277. https://doi.org/10.1016/j.joi.2010.01.002
17. Hutchins BI, Yuan X, Anderson JM, Santangelo GM. Relative citation ratio (RCR): A new metric that uses citation rates to measure influence at the article level. PLoS Biol. 2016;14(9), e1002541, 25 pages. https://doi.org/10.1371/journal.pbio.1002541
18. Schubert A, Braun T. Relative indicators and relational charts for comparative assessment of publication output and citation impact. Scientometrics. 1986;9(5–6):281–291. https://doi.org/10.1007/BF02017249
19. Meyer M, Bhattacharya C. Commonalities and differences between scholarly and technical collaboration. Scientometrics. 2004;61(3):443–456. https://doi.org/10.1023/B:SCIE.0000045120.04489.80
20. Pouris A, Ho Y-S. Research emphasis and collaboration in Africa. Scientometrics. 2014;98(3):2169–2184. https://doi.org/10.1007/s11192-013-1156-8
21. Kahn M. Co-authorship as a proxy for collaboration: A cautionary tale. Sci Public Policy. 2018;45(1):117–123. https://doi.org/10.1093/scipol/scx052
22. Sooryamoorthy R. Transforming science in South Africa: Development, collaboration and productivity. Basingstoke: Palgrave Macmillan; 2015: p. 101–134. https://doi.org/10.1057/9781137493071
23. Wagner CS, Whetsell TA, Mukherjee S. International research collaboration: Novelty, conventionality, and atypicality in knowledge recombination. Res Policy. 2019;48:1260–1270. https://doi.org/10.1016/j.respol.2019.01.002

AUTHORS:
Radhamany Sooryamoorthy[1]
Mduduzi N.G. Mtshali[1]

AFFILIATION:
[1]Sociology Programme, School of Social Sciences, University of KwaZulu-Natal, Durban, South Africa

CORRESPONDENCE TO:
Radhamany Sooryamoorthy

EMAIL:
sooryamoorthyr@ukzn.ac.za

DATES:
Received: 21 Aug. 2019
Revised: 28 Nov. 2019
Accepted: 21 Dec. 2019
Published: 26 Mar. 2020

HOW TO CITE:
Sooryamoorthy R, Mtshali MNG. Research productivity in science and its relationship to race in South Africa. S Afr J Sci. 2020;116(3/4), Art. #6838, 6 pages. https://doi.org/10.17159/sajs.2020/6838

ARTICLE INCLUDES:
☒ Peer review
☐ Supplementary material

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

EDITORS:
Jane Carruthers
Hester du Plessis

KEYWORDS:
publications, transformation, collaboration, higher education

FUNDING:
National Research Foundation (South Africa)

Research productivity in science and its relationship to race in South Africa

The research productivity of scholars is ordinarily affected by a combination of factors such as their age, gender, academic age, rank, qualification, experience, discipline, collaboration and co-authorship. A factor not often included in the analysis of research productivity is race. We examined the inter-relationship between race and other pertinent variables of research productivity of scholars in South Africa, drawing data from two waves of study. We found that there was an increase in the proportion of African respondents and in the productivity of both African and Indian respondents. Compared to 2008, African respondents had higher mean values than the rest in measures such as papers written in the last year, papers published in foreign journals, combined measure of journal publications and in total productivity in 2014. A significant proportion of African respondents has been moved to South Africa.

Significance:

- The study reveals the complexities of research productivity of scholars in higher education and research institutions in South Africa and offers insights into the influence of race.
- The findings contribute to the study of transformation that is underway in higher education and research institutions in the country.

Introduction

For some time now, science and scientific research have been gaining increased attention, interest and recognition in South Africa. One of the measures to understand the growth, development and trends in scientific research is to study the research productivity of scientists in higher learning and research institutions where most of the research activities are being undertaken.

A few studies have specifically examined research productivity in South Africa. Matthews[1] for instance, focused on the relationship between the productivity (the number of papers and author's share) and rank of academics in South African universities. North et al.[2] explored demographic and academic factors in research productivity.

A number of factors are responsible for research productivity. Age, gender, qualification, academic experience and rank are some of them. Race is another variable that determines varying levels of research productivity. However, race as an influencing factor in research productivity has not been studied sufficiently. The analysis done by Eagan and Garvey[3] revealed that race and ethnicity are associated with differences in research productivity. Sooryamoorthy's[4] study of South African scholars also showed how research activities are differentially influenced by race.

We undertook this empirical study to understand the relationship between race and the research productivity of academics and researchers in South Africa. We specifically focused on publication productivity. This study was conducted in two waves in 2008 and 2014 and was driven by the following research questions:

1. Has there been a change in the publication productivity of academics and researchers in universities and research institutions in South Africa between the two periods of study?
2. What relationship between publication productivity and race is evident in South Africa? How are the African respondents different from the other racial groups in their publication productivity in higher learning and research institutions?
3. Is there any evidence in the data to suggest that transformation in terms of race has occurred in the higher learning and research institutions in the country?

Background of the study

Research productivity has been defined and studied widely in varied contexts. A number of measures prominently appear in the literature, including publications (peer-reviewed articles, chapters, and authored and edited books), citations, *h*-index, fields or disciplines[1,3,5,6], qualifications[2,6], age[6,7], gender[3,8-10], academic rank[1,11], study abroad[12], collaboration or co-authorship[11,13-16], career path[17,18], and so on.

Data and methods

The analysis presented covers publications in peer-reviewed journals (both national and foreign origin), and chapters in edited volumes and books (edited and monographs) that were published by the respondents during the 5 years prior to the date of the interview. The measure also included the number of papers written in the previous year. Specific periods (for instance, between 2003 and 2007, and between 2000 and 2014) were considered so that the respondents could provide accurate information. From these individual measures, total publication productivity was computed as the dependent variable. Relying on the review of literature, relevant independent variables were identified: race, rank, age, academic age, gender, education, field of study, immigration, time spent on doing research (in hours per an average day), years spent outside the country for higher education, professional meetings attended (in the country and overseas), and collaboration (collaborated years and partners in the career). These independent variables were tested for their significance in publication productivity before they were finally adopted for correlational and regression analyses.

Two waves of data were used. One was collected in 2008 and the other in 2014. In both waves, sample respondents were drawn from two higher learning institutions and four research institutes in the province of KwaZulu-Natal in South Africa. There were 22 science departments (in the disciplines of agricultural science, engineering science, life sciences, natural sciences and medical sciences) in the selected institutions. As we intended to have as many respondents as possible, all those who were employed in these departments were approached for a face-to-face interview, using a structured interview schedule. Race and gender were self-described by respondents. Ethical clearance was obtained from the University of KwaZulu-Natal (ref. HSS/0336/014).

Analysis and findings

Background of the respondents

The demographic and professional details of the sample respondents are presented in Table 1. A total of 204 academics and researchers were interviewed in 2008 and 113 in 2014. Respondents were sampled from two sectors namely, academic and research institutions. Academics formed 69% of the sample and the remaining respondents were researchers. In this proportion no significant association between the samples for 2008 and 2014 was observed, three-fourths of the respondents belonged to the university sector and the remaining to the research institute sector.

The majority of respondents (79%) were in senior ranks such as senior lecturer (23%), associate or full professor (35%), and senior researcher/scientist (21%). The year-wise breakdown showed association between junior ranks and 2008, except for the rank of senior researcher/scientist.

The greatest proportion of respondents (43%) was from the natural sciences, followed by the life sciences (28%) and engineering (16%). Only a few came from agriculture and medicine. A difference in association between fields and the year of survey was evident. There were more respondents who worked in the natural sciences in 2008 than in 2014, and more engineering respondents in 2014 than in 2008. One-third of the total sample obtained degrees at overseas institutions.

About three-fourths of respondents were men and one-fourth women. The men:women ratio was 70:30 in 2008 and 80:20 in 2014. Two-thirds of the respondents in the whole sample were married and 28% were single at the time of the interview. Between the two waves, these percentages did not show any significant association. The mean age of the respondents for the entire sample was 44 years, which differed significantly between 2008 (42 years) and 2014 (47 years), as shown in the independent *t*-test. The career age of the respondents was about 11 years. The same was observed for both sets of respondents. The work experience of the respondents in the institutions where they worked at the time of the survey was less than their career age. This figure was about 10 years for both the whole sample and the sub-samples for 2008 and 2014.

Four major racial groups were identified in the sample: Africans, Indians, whites and coloureds. In South Africa in 2011, the population percentages were 79.5% African, 9% white, 9% coloured and 2.5% Indian.[19] In the total sample, the majority were white (49%), followed by African (26%) and Indian (22%). Coloured respondents, who also included some Asians, were in the minority (4%). The percentages of racial representation in the two periods did show a statistically significant difference of association in the chi-square test. The proportion of the African respondents increased from 20% in 2008 to 35% in 2014. Meanwhile the percentage of white respondents had decreased from 53 in 2008 to 40 in 2014. The same pattern was observed for the Indian respondents (24% in 2008 to 20% in 2014). The details of other variables are shown in Table 1.

Respondents and race

Almost two-fifths (39%) of the African respondents in 2008 were lecturers, as were 46% of Indians, but only 17% of whites. The percentage of respondents in the rank of senior lecturer was higher for Africans than for whites and Indians. In the first wave of the study, there were fewer associate professors and no professors among the African respondents. This was also the case for senior researcher/scientists among this racial group. By 2014, the proportion of African respondents in the various ranks had improved: there were more senior lecturers, associate professors and professors who were Africans. This finding is comparable with those for Indians and whites. However, representation in the rank of senior researcher/scientist was lower for African respondents than for Indian and white respondents in 2014.

Table 1: Respondents surveyed, 2008 and 2014

Respondents' details	2008		2014		Total	
	No.	%	No.	%	No.	%
Rank*						
Lecturer	60	29.4	0	0.0	60	18.9
Senior Lecturer	32	15.7	42	37.1	74	23.3
Associate Professor	13	6.4	25	22.1	38	12.0
Professor	22	10.8	16	14.2	38	12.0
Junior Researcher/Scientist	0	0.0	13	11.5	13	4.1
Senior Researcher/Scientist	51	25.0	14	12.4	65	20.5
Other Researcher/Scientist	16	7.8	3	2.7	19	6.0
Other Academics	10	4.9	0	0.0	10	3.2
Total	204	100	113	100	317	100
Job category						
Academic	137	67.2	83	73.5	220	69.4
Research	67	32.8	30	26.5	97	30.6
Organisational sector*						
University/Academic	141	69.1	98	86.7	239	75.4
Research institute	63	30.9	15	13.3	78	24.6
Born in South Africa*	**142**	**69.6**	**60**	**53.1**	**202**	**63.7**
Previously worked						
University	91	45.5	47	41.6	138	44.1
Research institute	23	11.5	23	20.4	45	14.7
Private sector	17	8.5	10	8.8	27	8.6
NGO/other	32	16.0	11	9.7	43	13.7
No prior employment	37	18.5	22	19.5	59	18.8
Highest degree*						
PhD	108	53.2	83	73.5	191	60.4
Masters	54	26.6	18	15.9	72	22.8
Bachelor	22	10.8	11	9.7	33	10.4
Diploma/other	19	9.4	1	0.9	20	6.3
Field of study*						
Natural sciences	95	50.5	33	29.2	128	42.5
Agriculture	15	8.0	10	8.8	25	8.3
Life sciences	58	30.9	27	23.9	85	28.2
Engineering	10	5.3	39	34.5	49	16.3
Medicine	10	5.3	4	3.5	14	4.7
Degree from overseas	**51**	**29.1**	**32**	**30.2**	**83**	**29.5**
Race**						
African	41	20.1	40	35.4	81	25.6
Indian	48	23.5	22	19.5	70	22.1
White	109	53.4	45	39.8	154	48.6
Other	6	2.9	6	5.3	12	3.8
Gender***						
Men	142	69.6	90	79.6	232	73.2
Women	62	30.4	23	20.4	85	26.8
Marital status						
Married	130	63.7	82	72.6	212	66.9
Single	61	29.9	29	25.7	90	28.4
Other	13	6.4	2	1.8	15	4.7
Age#	**42.1**	**10.8**	**46.8**	**12.0**	**43.8**	**11.5**
Work experience in the current institution	**10.8**	**10.4**	**9.76**	**9.7**	**10.4**	**10.1**
Career (academic age)	**11.8**	**9.6**	**10.9**	**10.1**	**11.5**	**9.7**

Chi-square test: *p=<*0.01*, **p=<*0.05*, ***p=<*0.1*
Independent t*-test:* #p=<*0.001*

The association between academics and racial groups was statistically significant based on the chi-square test. At the same time, there was an increase in the percentage of African academics (from 68% in 2008 to 88% in 2014) but not in that of Indian and white academics in 2014. In 2008, 50% of the African respondents were born in South Africa compared with 94% of Indians, 67% of whites and 70% of all respondents. By 2014 the percentage had decreased significantly for Africans (23%), than for Indians (77%) and whites (71%). In the sub-sample of 2014, the overall percentage was 53. This difference is large for Africans. This information should be considered in conjunction with the years of being in the country after they had moved to South Africa. In 2008, the average years of stay in the country for the whole sample was 17 years. Amongst the different racial categories, the average years of stay were 5.6 for Africans, 3.7 for Indians, 25 for whites and 3.8 for 'Others'. Africans and Indians (and Others) came to the country quite recently. In 2014, the figures were 10.5 years for Africans, 5.8 for Indians, 14.2 for whites and 7.7 for Others. The average for the population was 13 years. Here again both Indians and Africans were the most recently migrated respondents. For both surveyed years the difference was statistically significant.

In the 2008 sample, 50% of the African respondents had previous experience of working in a university and 13% in research institutes. The white respondents (64%) had prior working experience in similar work environments of universities and research institutes. The combined percentage for prior experience in universities and research institutes was higher (75%) for the African respondents in 2014 and higher than those of Indians (44%) and whites (58%). Compared to whites and Indians, fewer Africans had a PhD at the time of the survey in 2008. About 66% of the African respondents were from the field of natural sciences followed by 20% who worked in the field of life sciences. This finding can be compared with a similar percentage for the Indian respondents in the natural sciences and in the life sciences and a different percentage for the white respondents (37% natural sciences and 42% life sciences). By 2014, there was a similar percentage (30% each) of Africans in the fields of the natural sciences, life sciences and engineering. A diverse pattern emerged for the Indian and the white respondents. For most of the Indians the preferred fields were engineering (55%) and the natural sciences (36%). A more or less even spread across the fields of the natural sciences, agriculture, life sciences and engineering was the observed pattern for the whites. About one-third of the African respondents in the sample for 2008, which was similar to the percentage for the whole sample for 2008, had a degree earned overseas. This proportion is higher than that of the Indians (13%) but lower than that of the whites (37%). By 2014, the proportion had changed to 56% for Africans, 5% for Indians and 23% for whites, while the sample as a whole remained at 30%.

The ratio of men and women among the African sample for the year 2008 was similar to that of the entire sample. Men formed more than two-thirds of the African sample, and the percentage was not different from those for Indians and whites. The mean age of the African respondents was 5 years less than that of the 2008 sample. They were also younger than the Indians (by 4 years), and the whites (by 8 years). In the 2014 sample, the average age of the Africans was close to the average age for the sample, higher than that of the Indians by 9 years, and about the same as that of the white respondents. In regard to the work experience in their present institutions, the African respondents had one of the lowest figures in the 2008 sample: 59% lower than the average for the sample, 61% lower than the Indians and 66% lower than the whites. In the 2014 sample, this gap closed for the sample (28% lower), the Indians (17% lower) and the whites (47% lower). In academic age, Africans were the junior ones among all the groups in the 2008 sample: an average of 6.8 years compared with 8.9 for Indians and 15.2 for whites. Clearly, Africans had far less academic experience than their counterparts. In the 2014 sample, the African respondents had a similar academic age as the sample, higher than that of Indians by 3 years and lower than that of whites by 4 years. This means Africans had a higher academic age in 2014 than in 2008.

Publication productivity and race

The productivity variables chosen for the analysis were papers written in the last year and work published during the previous 5 years. The publications included in the analysis were papers published in national and foreign journals, chapters in edited volumes, edited books and monographs. Because the focus of the study was on peer-reviewed academic publications, those written in English were analysed. From these variables a combined publication productivity was computed. These variables were examined across the racial categories of Africans, Indians, whites and Others using an analysis of variance (ANOVA) test.

Before examining racial differences, the publication productivity of the samples across the two years, 2008 and 2014, were examined. These data are shown in Table 2. ANOVA results indicate that there were significant differences between 2008 and 2014 in the number of papers produced in the previous year of the survey, papers in foreign journals, papers in both foreign and national journals combined, edited books, and in the combined productivity. In all these variables the mean was higher for 2014 than for 2008. Statistically insignificant differences between 2008 and 2014 were found in the number of papers in national journals and in chapters in books.

The publication productivity of the respondents across races is presented separately for the samples for 2008 and 2009 in Tables 3 and 4. In the number of papers written in the previous year of the survey, the African respondents reported a mean value below that of the sample for 2008. They also wrote fewer papers in comparison to their white colleagues but slightly more than their Indian colleagues (Table 3). Peer-reviewed papers were published more in foreign journals than in national journals. African academics published 3 papers in foreign journals, which was more than the number the Indian respondents published (by 0.5 paper), fewer than the white respondents (by 4 papers) and less than all respondents in the sample for 2008 (by 2 papers). In the number of papers published in national journals, the African respondents also lagged behind the sample and the white respondents, while they were on a par with the Indian respondents. When the papers in national and foreign journals were combined, Africans published fewer papers than the sample overall for 2008 by 3 papers and than whites by 5.8 papers. However, they published 0.3 papers more than their Indian peers.

The combined measure of all publication variables provides an overall picture of the full publication productivity of the respondents. In this measure for 2008, the African respondents were below the average for the combined sample and the white respondents (Table 3). The mean

Table 2: Publication productivity in 2008 and 2014

Publication variables	2008		2014		Total		ANOVA (d.f.=1)	
	Mean	s.d.	Mean	s.d.	Mean	s.d.	F	*p*-value
Papers written in the last year	3.3	3.6	4.8	8.2	3.9	5.7	4.594	0.33
Papers in foreign journals	5.0	15.4	10.7	20.7	7.0	17.7	7.585	0.006
Papers in national journals	1.8	7.0	1.5	2.6	1.7	5.8	0.1226	0.635
Combined count of papers in journals	6.9	21.6	12.1	21.7	8.7	21.7	4.326	0.038
Chapters in books	0.7	2.3	0.5	1.7	0.6	2.1	1.134	0.288
Edited books	0.2	0.5	0.7	1.8	0.4	1.2	15.596	0.000
Monographs/books	0.1	0.4	0.2	0.6	0.1	0.5	2.311	0.129
Combined productivity of all publications	7.8	22.9	13.5	23.9	9.8	23.4	4.264	0.040

figure for Africans was 4.5 against 7.8 for the sample, 10.9 for the white respondents, and 3.8 for the Indians. This means they produced 43% less than an average respondent for the year, 59% less than the white respondents and 16% more than the Indian respondents.

The African respondents in the 2014 sample did well in the measures of productivity (Table 4). They wrote more papers than the sample as a whole (1.4 more), and than the Indian (1.8 more) and white (0.6 more) respondents. They published more papers in foreign journals than the sample (4.4 papers more), Indian (7.5 papers more) and white colleagues (3 papers more). In national journals, Africans published more than the sample and Indians (by 0.1 and 0.3 papers, respectively). In comparison with whites, the percentage was 0.1% lower for the Africans. They also published more chapters in edited volumes than the sample and the white respondents but about the same as the Indians. Finally, in the total publication productivity, African academics and researchers had a higher score than the sample mean, whites and Indians: about 83% over the Indians, 55% over the whites and 74% over the sample mean.

As seen in the analysis, there has been a substantial increase in the productivity of African and Indian respondents from 2008 to 2014. One reason for this increase can be attributed to the concerted efforts by the institutions to encourage them to develop academically, which forms part of the national policy to support previously disadvantaged groups. On the other hand, the small percentage of African respondents who were born in South Africa might raise questions about the intended or unintended consequences of the BEE (Black Economic Empowerment) policy of the government.

In order to determine the significant differences between the African respondents and other groups in the measures of publication productivity, we ran an independent *t*-test. Significant differences in the two-tailed *t*-test were obvious in the number of papers written in the previous year, papers in both national and foreign journals, combined measure for journal publications, edited books and in the total productivity. In all these measures, the white respondents reported higher mean values than did the Africans. In 2014, there were three measures that showed significant mean differences between Africans and whites: papers in foreign journals, the combined count of papers in journals, and total productivity of all publications. In all these three measures, Africans had a higher average than whites. No statistically significant difference was observed between Africans and Indians for any measure. This picture, however, changed in 2014: African respondents reported higher mean values than the other groups in four measures – papers written in the last year, papers published in foreign journals, combined count of journal papers, and combined total productivity.

The increase in the number of publications in foreign journals by African respondents, and to some extent Indian respondents, could be for two reasons. The first is the possibility of previous exposure to foreign journals by being born outside South Africa, and the second is the effect of overseas-based education and obtaining a degree overseas.

Publication productivity and influencing variables

The total publication productivity of the African respondents in 2008 was significantly and positively correlated with age, having a PhD, the number of years spent outside the country for educational and professional purposes, the number of collaborated years in career, and the number of collaborated partners in career (Table 5). Along with this, the number of publications in national and foreign journals and the number of co-published papers were significantly associated with total publication productivity. In 2014, the total publication productivity of Africans was influenced by the number of publications in foreign journals, co-published papers, having a PhD, and collaborated years in their career. In comparison to the African respondents, both Indian and white respondents showed similar correlations between the total publication productivity with papers in national and foreign journals and co-published papers. In the case of the white respondents for 2008, age, academic age, PhD and the time spent on doing research were important. These variables were not significant for the 2014 sample. The number of professional meetings attended overseas, collaborated years and collaborated partners also had a positive effect on the total productivity of white respondents in 2008, but only collaborated partners had an association in 2014. Among the Indian respondents in 2008, three factors – namely having a PhD, the number of overseas professional meetings attended and collaborated years – were significant along with papers in national and foreign journals and co-published papers. In 2014, the total publication productivity of the Indian respondents was positively affected by papers in foreign journals, co-published papers, the number of overseas professional meetings attended in the previous 5 years and collaborated partners in their career.

Productivity is affected by rank, age, academic age, gender, marital status, race, highest degree, exposure to overseas education, and collaboration.

Table 3: Publication productivity across racial categories in 2008

Publication variables	African		Indian		White		Other		Total		ANOVA (d.f.=3)	
	Mean	s.d.	Mean	s.d.	Mean	s.d.	Mean	s.d.	Mean	s.d.	F	*p*-value
Papers written in the last year	2.6	2.4	2.3	2.1	4.1	4.3	2.3	1.6	3.3	3.6	3.941	0.009
Papers in foreign journals	3.0	5.1	2.5	5.5	7.0	20.4	3.3	5.3	5.0	15.4	1.291	0.279
Papers in national journals	0.9	1.9	0.9	2.3	2.7	9.3	0.8	1.3	1.8	7.0	1.126	0.340
Combined count of papers in journals	3.9	5.8	3.3	5.8	9.7	28.8	4.2	5.0	6.9	21.6	1.335	0.264
Chapters in books	0.5	1.3	0.3	0.7	1.0	2.9	0.7	1.2	0.7	2.3	1.109	0.346
Edited books	0.1	0.3	0.1	0.4	0.2	0.5	0.5	0.8	0.2	0.5	1.695	0.169
Monographs/books	0.1	0.2	0.0	0.2	0.1	0.4	0.7	1.2	0.1	0.4	4.699	0.003
Combined productivity of all publications	4.5	6.4	3.8	6.1	10.9	30.5	6.0	6.9	7.8	22.9	1.472	0.223

Table 4: Publication productivity across racial categories in 2014

Publication variables	African		Indian		White		Other		Total		ANOVA (d.f.=3)	
	Mean	s.d.	Mean	s.d.	Mean	s.d.	Mean	s.d.	Mean	s.d.	F	*p*-value
Papers written in the last year	5.4	9.0	3.6	3.4	4.8	9.2	5.3	6.8	4.8	8.2	0.212	0.888
Papers in foreign journals	15.1	27.7	7.6	9.6	8.7	18.1	7.5	7.7	10.7	20.7	0.945	0.422
Papers in national journals	1.6	2.7	1.3	1.8	1.7	3.0	0.0	0.0	1.5	2.6	0.834	0.478
Combined count of papers in journals	16.7	28.2	8.9	9.5	10.3	20.2	7.5	7.7	12.1	21.7	0.947	0.421
Chapters in books	0.6	2.4	0.6	1.7	0.3	0.9	0.0	0.0	0.5	1.7	0.448	0.719
Edited books	0.7	2.4	0.2	0.5	1.0	1.6	0.3	0.5	0.7	1.8	0.884	0.452
Monographs/books	0.2	0.6	0.2	0.6	0.2	0.6	0.0	0.0	0.2	0.6	0.210	0.889
Combined productivity of all publications	18.2	32.5	9.9	10.5	11.7	20.4	7.8	7.5	13.5	23.9	0.883	0.452

In order to explore the influence of these factors we ran regression of productivity on the selected control variables. Multicolinearity was checked for highly correlated variables. Three models with their standardised beta coefficients and significance are presented in Table 6. The first model on the productivity of publications in foreign journals indicates that two factors – race and collaborated years – are positively correlated. The model explains 19% of variance in the productivity of publications in foreign journals. In agreement with the previous findings, African respondents were correlated with their publication productivity in foreign journals. It is also evident that the number of collaborated years in a person's career is positively correlated with their publications in foreign journals. The second and third models relating to the productivity of co-publications and total productivity also showed significant correlation of the same two variables (race and collaborated years), and explains 18% of variance.

Discussion

The objective of this study was to analyse the relationship between the publication productivity and racial categories of academics and researchers in South African higher learning and research institutions. In particular, it was our intention to examine how transformation in the academic and research environment has occurred in South Africa. It should be noted that the sample was derived from only one of the provinces in South Africa and therefore the generalisability of the findings are limited.

The proportion of respondents in the sample did not correspond with the national demographic composition, in which Africans are the majority with 79.5% of the total population.[19] This finding has been confirmed by several authors, including Govinder et al.[20] and Habib and Morrow[21]. Staffing in the premier institutions in the country, as Soudien[22] reported, remains largely white. Transformation in higher education is part of a broader process of South Africa's political, social and economic transition.[23]

There has been a positive change in the proportion of African respondents, particularly in higher-ranking positions, between 2008 and 2014. By 2014, the publication productivity of African respondents had changed for the better and in total productivity they performed better than the sample as a whole and the other racial categories. In a recent analysis, Mouton et al.[24] reported that the proportion of papers by black authors increased substantially from 16% in 2005 to 31% in 2016, which is in line with our findings.

Conclusion

The analysis presented in this paper brings forth issues that relate to productivity and race in South Africa. Notable and significant changes in the publication productivity of scholars were evident in the two waves of data. Racial transformation is key for developmental goals, but is not occurring based on our sample. This has implications for the

Table 5: Correlations between publication productivity and other relevant variables, 2008 and 2014

Variables	Total publication productivity					
	African		Indian		White	
	2008	2014	2008	2014	2008	2014
Papers in national journals	0.494 **	0.213	0.379 **	0.093	0.928 **	0.720 **
Papers in foreign journals	0.923 **	0.995 **	0.884 **	0.966 **	0.985 **	0.987 **
Co-published papers in journals	0.807 **	0.982 **	0.920 **	0.912 **	0.979 **	0.981 **
Age	0.440 **	-0.098	0.193	0.315	0.235 *	0.079
Academic age	0.227	0.141	0.257	0.178	0.265 **	0.141
Having a PhD	0.643 **	0.339 *	0.513 **	0.421	0.224 *	0.256
Hours spent on research	-0.170	0.190	0.129	-0.119	0.213 *	0.204
Years spent overseas for higher education	0.467 *	0.084	0.222	0.536	0.108	0.053
Professional meetings attended	0.048	0.161	-0.127	0.366	0.010	-0.095
Professional meetings in country	0.185	0.244	0.182	0.339	0.074	0.007
Professional meetings overseas	0.134	0.229	0.479 **	0.479 *	0.818 **	0.110
Collaborated years in career	0.454 **	0.411 *	0.491 **	0.684 **	0.263 **	0.226
Collaborated partners in career	0.444 *	0.035	-0.109	0.423	0.307 **	0.445 **

**p=<*0.01*, *p=<*0.05*

Table 6: Regression of productivity on factors

Control variables	Productivity in foreign journals		Productivity in co-publications		Total productivity	
	Standardised beta	*p*-value	Standardised beta	*p*-value	Standardised beta	*p*-value
Sector (1=academia, 0=others)	0.065	0.481	0.077	0.405	0.071	0.444
Rank (1=seniors, 0=others)	0.087	0.331	0.069	0.445	0.074	0.409
Age	-0.180	0.308	-0.156	0.378	-0.132	0.457
Academic age	0.146	0.833	0.162	0.361	0.167	0.346
Gender (1=male, 0=others)	0.013	0.141	0.000	0.999	-0.010	0.910
Race (1=African, 0=others)	0.179	00.65	0.190	0.051	0.171	0.079
Marital status (1=married, 0=others)	0.132	0.155	0.122	0.191	0.099	0.288
Born in South Africa (1=yes, 0=no)	0.049	0.658	0.072	0.513	0.082	0.455
Highest degree (1=PhD, 0=others)	0.148	0.151	0.143	0.167	0.131	0.204
Degree from developed countries (1=yes, 0=no)	-0.082	0.446	-0.072	0.506	-0.105	0.335
Years spent outside South Africa for higher studies	0.118	0.447	0.169	0.280	0.117	0.455
Years spent in developed countries	-0.031	0.839	-0.070	0.649	-0.015	0.920
Collaborated years in the entire career	0.260	0.038	0.236	0.060	0.242	0.055
Collaborated people in the entire career	0.057	0.536	0.057	0.534	0.080	0.384
n	133		133		133	
R^2	0.186		0.181		0.178	

scientific personnel in the country and the efforts to address issues of skills shortage and capacity. The study showed that there were African respondents who had moved to South Africa to take up positions as academics and researchers. This, however, cannot be regarded as something that facilitates transformation. If that were the case, there should have been more native Africans and Indians than foreign nationals. We do not suggest that South Africa should be a closed society for foreign nationals, but it should focus more on its own people to achieve the ideals of transformation. One of the recommendations of the Soudien Committee was that there should be a GOOT (grow one's own timber) programme. This is pertinent to make transformation a reality and meaningful for both the higher education and research system and for the country.

Acknowledgement

We acknowledge funding received from the National Research Foundation (South Africa) to conduct this study.

Authors' contributions

R.S. conceptualised the study and collected the data. R.S. and M.N.G.M. analysed the data and wrote the draft and final versions.

References

1. Matthews AP. Physics publication productivity in South African universities. Scientometrics. 2013;95:69–86. https://doi.org/10.1007/s11192-012-0842-2
2. North D, Zewotir T, Murray M. Demographic and academic factors affecting research productivity at the University of KwaZulu-Natal. S Afr J High Educ. 2015;25:1416–1428.
3. Eagan Jr MK, Garvey JC. Stressing out: Connecting race, gender, and stress with faculty productivity. High Educ. 2015;86:924–954. https://doi.org/10.1353/jhe.2015.0034
4. Sooryamoorthy R. Internet technology and the epistemic strategies of scientists in post-apartheid South Africa: Race as a decisive factor. Tech Soc. 2015;40:73–81. https://doi.org/10.1016/j.techsoc.2014.08.003
5. Abramo G, D'Angelo CA. How do you define and measure research productivity? Scientometrics. 2014;101:1129–1144. https://doi.org/10.1007/s11192-014-1269-8
6. Prpić K. The publication productivity of young scientists: An empirical study. Scientometrics.2000;49:453–490.https://doi.org/10.1023/A:1010541707878
7. Bonaccorsi A, Daraio C. Age effects in scientific productivity: The case of the Italian National Research Council (CNR). Scientometrics. 2003;58:49–90. https://doi.org/10.1023/A:1025427507552
8. Van Arensbergen P, Van der Weijden I, Van den Besselaar P. Gender differences in scientific productivity: A persisting phenomenon? Scientometrics. 2012;93:857–868. https://doi.org/10.1007/s11192-012-0712-y
9. König CJ, Fell CB, Kellnhofer L, Schui G. Are there gender differences among researchers from industrial/organizational psychology? Scientometrics. 2015;105(3):1931–1952. https://doi.org/10.1007/s11192-015-1646-y
10. Mayer SJ, Rathmann JMK. How does research productivity relate to gender? Analyzing gender differences for multiple publication dimensions. Scientometrics. 2018;117:1663–1693. https://doi.org/10.1007/s11192-018-2933-1
11. Abramo G, D'Angelo CA, Costa FD. Research collaboration and productivity: Is there correlation? High Educ. 2009;57:155–171. https://doi.org/10.1007/s10734-008-9139-z
12. Shin JC, Jung J, Postiglione GA, Azman N. Research productivity of returnees from study abroad in Korea, Hong Kong, and Malaysia. Minerva. 2014;52:467–487. https://doi.org/10.1007/s11024-014-9259-9
13. Duque RB, Ynalvez M, Sooryamoorthy R, Mbatia P, Dzorgbo D-B, Shrum W. Collaboration paradox: Scientific productivity, the internet, and problems of research in developing areas. Soc Stud Sci. 2005;35:755–785. https://doi.org/10.1177/0306312705053048
14. Lee S, Bozeman B. The impact of research collaboration on scientific productivity. Soc Stud Sci. 2005;35:673–702. https://doi.org/10.1177/0306312705052359
15. Sooryamoorthy R. Publication productivity and collaboration of researchers in South Africa: New empirical evidence. Scientometrics. 2014;98:531–545. https://doi.org/10.1007/s11192-013-0990-z
16. Sooryamoorthy R. Transforming science in South Africa: Development, collaboration and productivity. Hampshire and New York: Palgrave Macmillan; 2015.
17. Fukuzawa N. An empirical analysis of the relationship between individual characteristics and research productivity. Scientometrics. 2014;99:785–809. https://doi.org/10.1007/s11192-013-1213-3
18. Lee DH. Predicting the research performance of early career scientists. Scientometrics. 2019;121:1481–1504. https://doi.org/10.1007/s11192-019-03232-7
19. Statistics South Africa. Mid-year population release 2011. Pretoria: Statistics South Africa; 2011.
20. Govinder KS, Zondo NP, Makgoba MW. A new look at demographic transformation for universities in South Africa. S Afr J Sci. 2013;109:86–96. http://dx.doi.org/10.1590/sajs.2013/20130163
21. Habib A, Morrow S. Research, research productivity and the state in South Africa. Transformation. 2006;62:9–29. https://doi.org/10.1353/trn.2007.0006
22. Soudien C. Transformation in higher education: A briefing paper. Johannesburg: Development Bank of Southern Africa; 2010.
23. South African Department of Education. Education White Paper 3: A programme for the transformation of higher education. Pretoria: Department of Education; 1997.
24. Mouton J, with Basson I, Blanckenber J, Boshoff N, Prozesky H, Redelinghuys H, et al. The state of the South African research enterprise. Stellenbosch: DST-NRF Centre of Excellence in Scientometrics and Science, Technology and Innovation Policy, Stellenbosch University; 2019.

AUTHORS:
Matthew J. McGill[1]
Robert J. Swap[1,2]
John E. Yorks[1]
Patrick A. Selmer[3]
Stuart J. Piketh[2]

AFFILIATIONS:
[1]NASA Goddard Space Flight Center, Greenbelt, Maryland, USA
[2]Department of Environmental Sciences and Management, North West University Potchefstroom, South Africa
[3]Science Systems and Applications Inc., Lanham, Maryland, USA

CORRESPONDENCE TO:
Matthew McGill

EMAIL:
matthew.j.mcgill@nasa.gov

DATES:
Received: 02 Oct. 2019
Revised: 13 Dec. 2019
Accepted: 21 Dec. 2019
Published: 26 Mar. 2020

HOW TO CITE:
McGill MJ, Swap RJ, Yorks JE, Selmer PA, Piketh SJ. Observation and quantification of aerosol outflow from southern Africa using spaceborne lidar. S Afr J Sci. 2020;116(3/4), Art. #6398, 6 pages. https://doi.org/10.17159/sajs.2020/6398

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:
lidar, aerosols, aerosol transport, biomass burning

FUNDING:
None

Observation and quantification of aerosol outflow from southern Africa using spaceborne lidar

Biomass burning in Africa provides a prolific source of aerosols that are transported from the source region to distant areas, as far away as South America and Australia. Models have long predicted the primary outflow and transport routes. Over time, field studies have validated the basic production and dynamics that underlie these transport patterns. In more recent years, the advancement of spaceborne active remote-sensing techniques has allowed for more detailed verification of the models and, importantly, verification of the vertical distribution of the aerosols in the transport regions, particularly with respect to westerly transport over the Atlantic Ocean. The Cloud-Aerosol Transport System (CATS) lidar on the International Space Station has detection sensitivity that provides observations that support long-held theories of aerosol transport from the African subcontinent over the remote Indian Ocean and as far downstream as Australia.

Significance:

- Biomass burning in Africa can have impacts as far away as Australia.
- Flow of aerosols from Africa towards Australia has long been postulated by transport models, but has been poorly characterised due to a lack of measurements.
- The CATS instrument on the International Space Station has detection sensitivity that captures aerosol transport from Africa over the Indian Ocean to Australia.

Introduction

The African continent is a prolific source of aerosols flowing out over the Atlantic and Indian Oceans. Transport of Saharan dust off the continent and over the equatorial and North Atlantic Ocean is well documented.[1-4] It is now appreciated that dust from the African subcontinent, following that transport route, finds its way to the Caribbean and Amazon basin.[5-10] Similarly, evidence of sub-Saharan aerosol and trace gas transports comprising biomass burning smoke, dust and industrial emissions has been documented. These transports fall into three general categories: (1) out over the Atlantic Ocean (originating primarily in tropical Africa north of 20ºS)[11-18]; (2) air mass recirculation from and over the southern portion of the subcontinent[19-21]; and (3) westerly transport out over the Indian Ocean (south of 20ºS)[13,14,18,22-33]. The transports and their emission sources mentioned above that contribute to the atmospheric aerosol loading over and off southern Africa exhibit a strong seasonality and tend to migrate from western tropical southern Africa in May to southeastern southern Africa and Mozambique in September and October.

The Southern African Fire-Atmosphere Research Initiative and Southern African Regional Science Initiative (SAFARI 92 and SAFARI 2000, respectively) were extensive field campaigns specifically designed to study the postulated aerosol and trace gas transports of combined emissions, in general, and biomass burning emissions, in particular, from the southern regions of the African continent.[31,34] While SAFARI 2000 focused on aerosol emissions and transports, it did so primarily over and very near the southern African subcontinent. Easterly and westerly atmospheric transports from southern Africa occur over expansive areas of the remote Atlantic and Indian Oceans where ground-based and sea surface measurements are sparse and airborne measurements are challenging to obtain. Understanding, following and documenting atmospheric features such as these, requires the use of atmospheric models and satellite data.

Near source regions, aerosol concentrations in outflows are dense and sufficiently optically thick to be rather easily detected by spaceborne passive sensors such as the Moderate Resolution Imaging Spectrometer (MODIS).[35-37] Although optically thick layers can be detected by MODIS or other passive sensors, over oceans, the aerosol optical depth cannot be accurately retrieved from MODIS for aerosol layers that have aerosol optical depths of less than 0.03.[36,37]

For less optically thick outflows, active remote sensors, such as the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO)[38] lidar, can be used to detect aerosol layers. However, CALIPSO also requires a minimum density of scatterers before an aerosol layer can be detected. A challenge to both passive and active sensors, as noted by Edwards et al.[39] is that high aerosol concentrations generally do not extend far from the source region. Far from the source region, aerosols are lofted and transported over the Indian Ocean. The aerosol plume tends to spread, somewhat in the horizontal but more in the vertical, thereby becoming too diffuse for spaceborne sensors to detect. This flow is in contrast to the easterly flow out over the Atlantic, which either occurs within the boundary layer, particularly over Namibia during offshore transport of surface dust[40], or is bounded between 800 hPa and 500 hPa[41].

Although transport models routinely predict aerosol plumes over Australia, measurements verifying the plume height and distribution are extremely limited. Some ground-based measurements from Australia have shown evidence of the outflow plume[26,28] but spaceborne measurements that can conclusively track the outflow from the source region to the Australian continent have been lacking. There were initial spaceborne lidar measurements made by the Laser In-space Technology Experiment (LITE)[42] that appear to capture a feature similar to those described in this paper during September 1994. As LITE was a technology demonstration onboard the Space Shuttle, those measurements were limited in coverage and, moreover, the 1064 nm data from LITE was never calibrated. In this paper, we present,

for the first time, calibrated 1064 nm observations that support long-held (>25 years) postulated understandings of atmospheric transport modes from the biomass burning region of subequatorial Africa out over the Indian Ocean and towards Australia.

The Cloud-Aerosol Transport System

There have been only two in-space lidar sensors that have operated over multiple years to capture seasonal transport patterns: CALIPSO and the Cloud-Aerosol Transport System (CATS) onboard the International Space Station (ISS). The CATS sensor is a backscatter lidar instrument with depolarisation measurement.[43] A notable feature of CATS is the use of photon-counting detection, which permits high detection sensitivity. As a result, at least during night portions of each orbit, CATS has detection sensitivity (minimum detectable backscatter, at 1064 nm) as low as 5 x 10^{-5} km^{-1} sr^{-1}, which is more sensitive than the CALIPSO minimum detection sensitivity (at 532 nm) of ~8 x 10^{-4} km^{-1} sr^{-1}.[44] As noted in the previous section, although the LITE demonstration had detection sensitivity sufficient to detect diffuse aerosol layers, the limited lifetime (approximately 40 h total observation time) and the limited number of observations over the study region precludes an ability to track individual events as they vary with synoptic conditions. The more continuous and multi-year operation of CATS, coupled with the high detection sensitivity, can be used to demonstrate persistence of the aerosol outflow as well as tracking of the outflow from Africa towards Australia.

Operating from February 2015 until October 2017, CATS data contains observations of each of the outflow patterns identified in Garstang et al.[13] The most intriguing are observations of the westerly transport of aerosols out over the Indian Ocean and over Australia. The detection sensitivity of CATS at 1064 nm has enabled observations of the diffuse aerosol plumes transported off the African subcontinent over those regions, providing direct measurement of the transport predicted by Garstang et al.[13] Moreover, unlike other spaceborne lidar sensors, the CATS 1064 nm data is directly calibrated at 1064 nm[44,45], thereby augmenting the available data record with additional wavelength information.

In addition to backscatter detection, the CATS 1064 nm channel provides a linear depolarisation measurement. The depolarisation measurement is exceptionally useful as an aid in determining cloud and aerosol type.[46,47] Relevant to African aerosol transport, where smoke and dust (and combinations of the two) are prevalent, the depolarisation ratio provides a critical determinant of aerosol type. Smoke tends to have a linear depolarisation ratio of the order of 1–10%, whereas dust is in the range of 20–30%. Smoke combined with dust will lower the ratio somewhat, typically to the 10–25% range. The depolarisation measurement provides important substantiation that the elevated layers observed are, in fact, composed of smoke particles and, hence, are coming from the expected source region. The LITE demonstration did not have depolarisation measurement capability, thus making those prior measurements more challenging to relate to aerosol type.

CATS observations of westerly outflow

The CATS lidar onboard the ISS, with its unique precessing orbit, has captured multiple occurrences of westerly outflow from the African subcontinent towards Australia. The CATS data used herein are calibrated Level 1B data products, specifically 1064 nm attenuated total backscatter coefficients at a resolution of 350 m horizontal by 60 m vertical.[44] Three specific examples are described below.

Case 1: 7 September 2016

Multiple ISS passes on 7 September 2016 provide a unique Eulerian perspective with multiple snapshots of the resultant transport plume. As illustrated in Figure 1, data captured on subsequent orbits show evolution of an aerosol plume originating off the west coast of southern Africa (approximate latitude 25°S) and propagating across the Indian Ocean to south of Australia. Analysis of 5-day back trajectories, shown in Figure 2a, obtained from the Hysplit model[48], indicates that parcels observed in the 9-km altitude range to the west of Australia on 7 September 2016 originated over the west coast of southern Africa (within the free troposphere) as well as from South America. Both these regions are large biomass burning source regions during the austral spring. The trajectory analysis illustrates how transport from the two continents merges in a transient westerly wave over southern Africa and exits the African subcontinent towards the southeast, south of the semi-permanent Indian anticyclone (Figure 3). In this transport pathway, air parcels rise rapidly and within 2 days are separated from the surface layer to become a clearly defined lofted layer. This observation is consistent with the postulated expectations based on modelled outputs (e.g. Garstang et al.[13], Tyson and D'Abreton[20]) that the westerly plumes exiting the subcontinent in a westerly wave tend to ascend over the southern Indian Ocean, facilitating rapid transport toward Australasia.

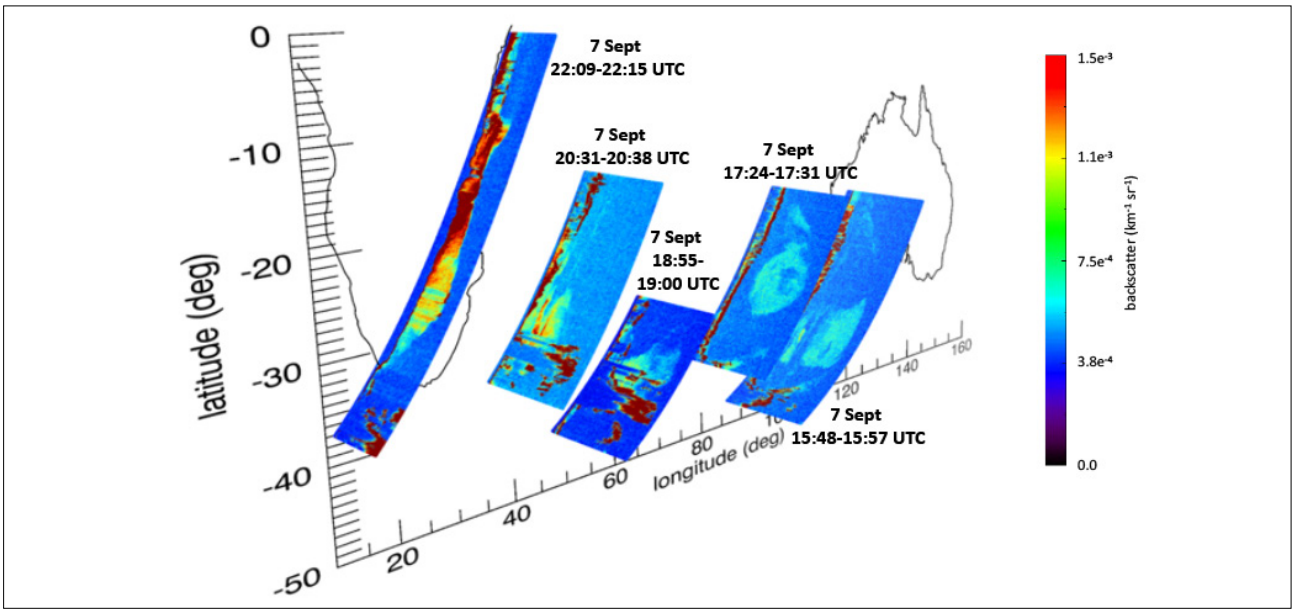


Figure 1: Case 1, 7 September 2016, showing an Eulerian view of plume transport. This figure visualises cross-sections of plumes on different passes during the same day. Time intervals of each cross-sectional segment are as noted on the image. The elevated plume is shown in light blue; the dark red/brown colour is cloud. The fourth track from the left (7 September 17:27 UTC) is further detailed in Figure 3.

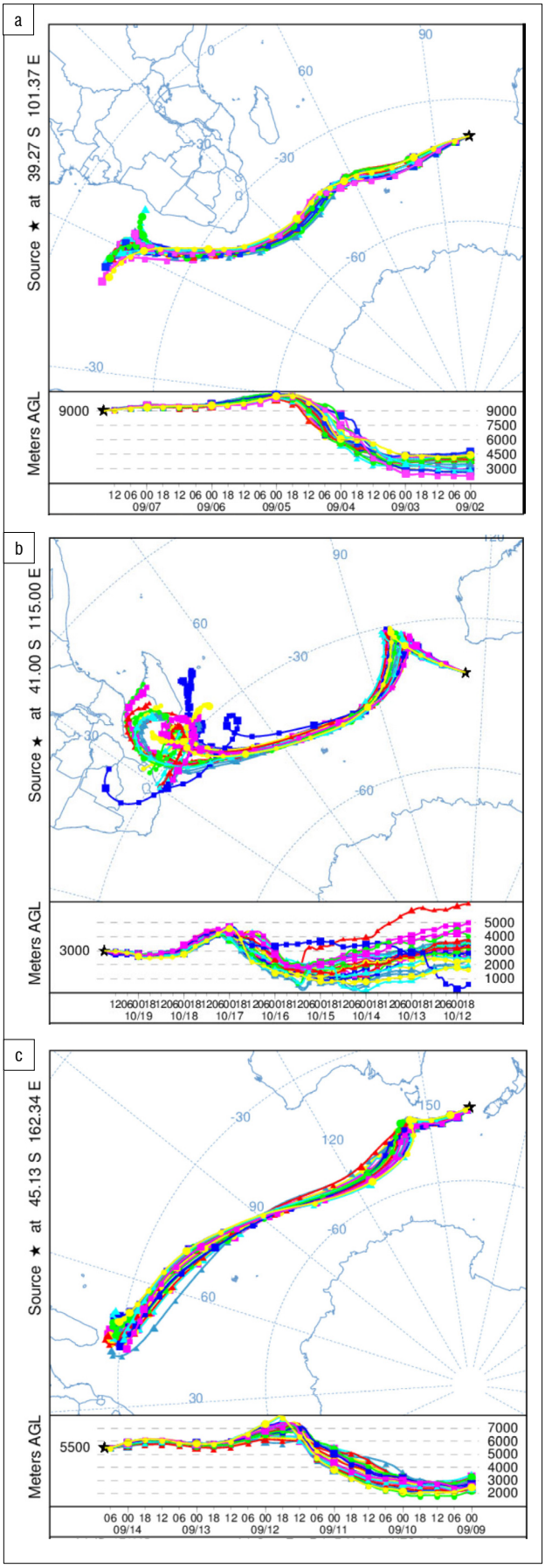
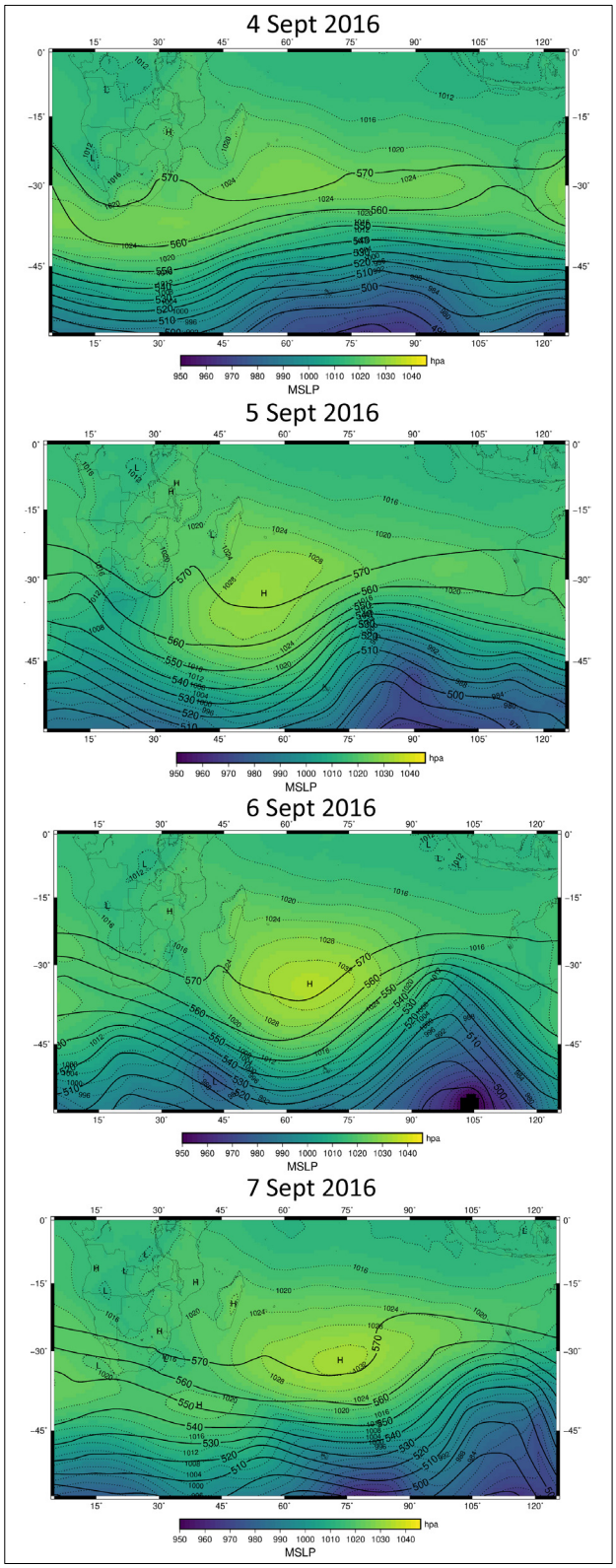


Figure 2: Back trajectory analyses from the NOAA Hysplit model for data shown in (a) Figure 1, 7 September 2016, (b) Figure 5, 11–18 October 2015 and (c) Figure 6, 13–19 September 2016.



MSLP, mean sea level pressure

Figure 3: European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis plots indicating the large-scale circulation from southern Africa to Australia for 4–7 September 2016 that coincides with Case 1. The dashed lines indicate the isobars at the surface while the solid lines give the isobars at 500 hPa. The shading is the geopotential height. The transient westerly wave, driving both the rapid atmospheric transport as well as the lofting mechanism, is seen clearly as it moves from the west of southern Africa and transitions to a location over the South Indian Ocean.

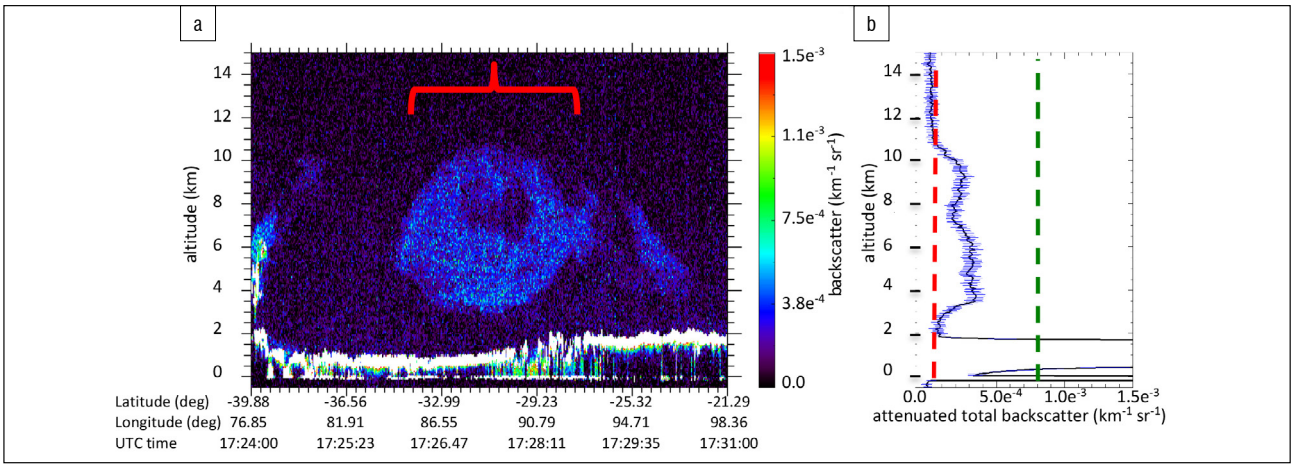


Figure 4: (a) Cross-section of the backscatter west of Australia on 7 September 2016. The profile on the right (b) is an average of backscatter through the layer (95-s average, as indicated by the red bracket on the figure) to show the structure and low magnitude (< 5 x 10^{-4} km^{-1} sr^{-1}) of the backscatter. The red dashed line on (b) indicates the CATS minimum detection threshold while the green dashed line indicates the same for CALIPSO.

Back trajectory analysis suggests the layer should be in the 9-km altitude range near Australia, and CATS profiles show the layer extending from an altitude of about 3 km up to about 11 km. This range in the vertical distribution is consistent with what Wenig et al.[30] found in the long-range transport of nitrogen dioxide plumes between southern Africa and Australia. Such transport is also consistent with previously described transports of water vapour[49], trace gases[50] and aerosols[51] off the subcontinent – transports that have been demonstrated to impact atmospheric chemistry and composition as well as possibly the biogeochemical cycling of precipitation and the ocean surface along the path of transport[52].

Figure 4 shows a cross-section of the elevated plume as it approaches the west coast of Australia. As seen in Figure 4b, the elevated layer is distinct and extends from 3 km up to about 11 km, but with low backscatter of <5 x 10^{-4} km^{-1} sr^{-1}. Although covering a large vertical extent, the median optical depth of the layer west of Australia is only of the order of 0.03–0.05 (±0.008). The median depolarisation ratio (integrated through the layer) is 0.05–0.08, indicating the elevated layer is composed primarily of smoke. The layer does start with higher optical depth (mean optical depth of 0.15±0.05) over the African subcontinent, which is attenuated as it is transported across the Indian Ocean through the loss of particles by wet and dry removal processes.[25]

Case 2: 11–18 October 2015

In contrast to the Eulerian view of Case 1, over the 8-day period of 11–18 October 2015, CATS captured a Lagrangian view of the evolution and transport of multiple plumes. Data captured during this period, displayed in Figure 5, show evolution of multiple dust or aerosol plumes that originate over southern Africa (approximate latitude 25°S) and propagate across the Indian Ocean to the south of Australia. Back trajectory analysis (Figure 2b) again indicates that 5–6 days are required for a plume to transit to Australia. Back trajectory analysis suggests that near the west coast of Australia the layer should be in the 3-km altitude range, and CATS profiles show the layer extending from an altitude of about 1 km up to about 8 km. The median optical depth of the layer west of Australia is of the order of 0.01–0.03 (±0.09). The median depolarisation ratio (integrated through the layer) is 0.05–0.10, again indicating primarily smoke.

Case 3: 13–19 September 2016

Similar to Case 2, this example presents a Lagrangian view of a plume transiting off southern Africa towards Australia. Figure 6 shows data captured during the period 13–19 September 2016, highlighting the evolution of a plume that transits directly over Australia and then continues on to the south. Similar to the other two cases, back trajectory analysis (Figure 2c) again indicates that 5–6 days are required for the plume to transit to Australia. Back trajectory analysis suggests the layer should be in the 5–6 km altitude range, and CATS profiles show the layer extending from an altitude as low as 1 km up to as high as 10 km. Although covering a large vertical extent, the median optical depth of the layer over Australia is low, only of the order of 0.008–0.01 (±0.003) and again becomes progressively lower the farther east the plume travels. The median depolarisation ratio (integrated through the layer) is 0.02–0.08, once again indicating that the layer is composed primarily of smoke.

Supporting meteorological information

Garstang et al.[13] showed that during the dry season in southern Africa (April through October), the dominating synoptic weather pattern is an anticyclonic circulation that results in horizontal recirculation at spatial scales as high as thousands of kilometres. Aerosols exit this anticyclonic flow in the southernmost part of Africa to the east into the Indian Ocean via westerly wave and trough disturbances. These westerly disturbances peak in the spring months (September–November) and in very dry seasons such as those observed during the SAFARI project in 1992, can direct as much as 90% of aerosol transport into the Indian Ocean.[13]

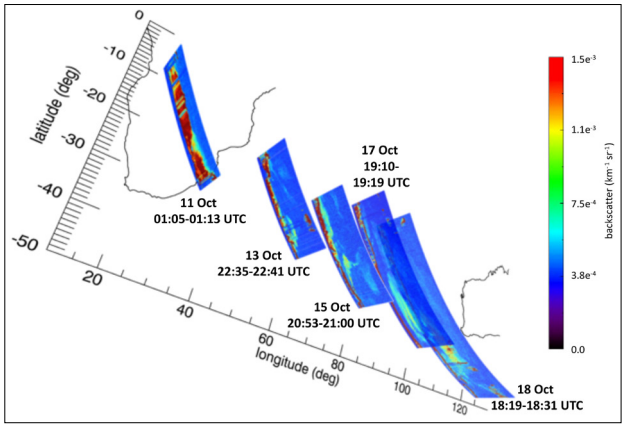


Figure 5: Case 2, 11–18 October 2015, demonstrating a Lagrangian view of plume transport, showing different passes on subsequent days. Time intervals of each cross-sectional segment are as noted on the image. The elevated plume is shown in light blue; the dark red/brown colour is cloud.

Synoptic weather maps of surface pressure and wind from the South African Weather Service were analysed for the three cases of smoke transport into the Indian Ocean observed by CATS. All three of these cases strongly support the observations made during SAFARI 92. Low pressure systems propagating across the southern edge of Africa, in tandem with high pressure located near Madagascar, result in flow towards the south and east that transports smoke into the Indian Ocean across 35°E. During this transport, the smoke is lofted and advected toward Australia in the

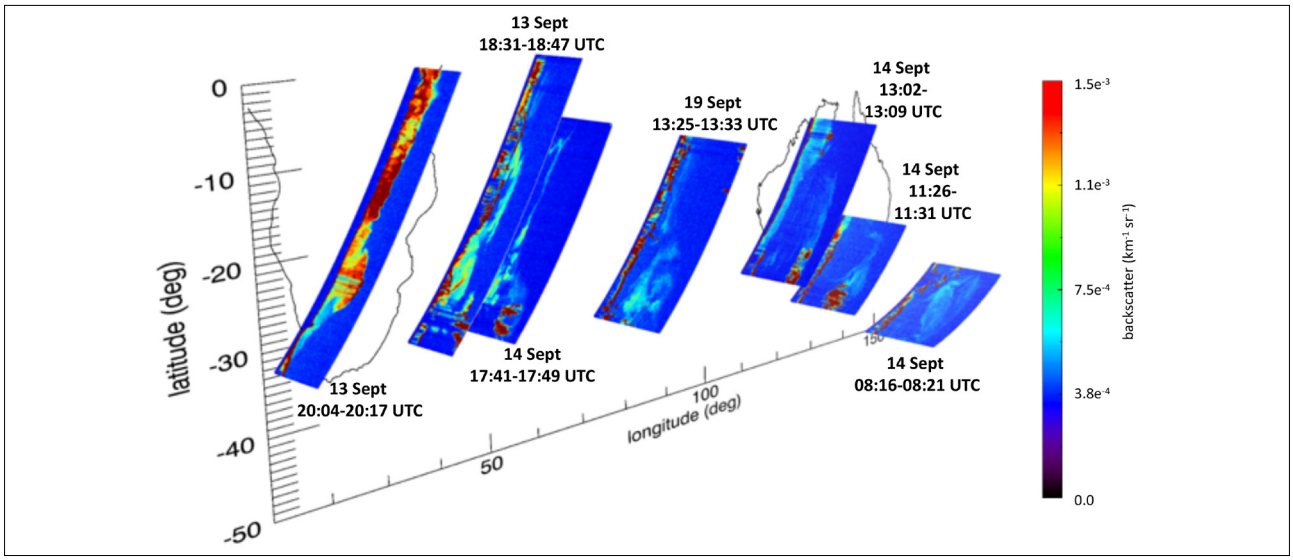


Figure 6: Case 3, 13–19 September 2016, demonstrating a Lagrangian view of plume transport, showing different passes on subsequent days. The elevated plume is shown in light blue; the dark red/brown colour is cloud.

prevailing westerlies at southern latitudes higher than 30°S. The global atmospheric circulation patterns were also similar between SAFARI 92 and the CATS cases in 2015 and 2016. The National Oceanographic and Atmospheric Administration (NOAA) Oceanic Nino Index, a rolling 3-month average of sea surface temperatures in the eastern tropical Pacific, indicates relatively strong El Nino conditions existed in 2015 and continued into most of 2016.

Back trajectories obtained from the Hysplit model confirm that transport from Africa to Australia generally took from 5 to 7 days. The back trajectories were initialised using the centroid of the elevated plume. In Cases 2 and 3, the trajectories trace back to the east coast of southern Africa (near the southern tip of Madagascar), whereas in Case 1 the trajectories trace to the west coast of southern Africa.

Conclusions

Biomass burning in Africa has long been recognised as a significantly important source of aerosols and trace gas. Focused field studies, such as SAFARI 92 and SAFARI 2000, validated the primary source and outflow patterns for smoke and trace gases from the African subcontinent. The primary outflow and transport routes from Africa to the Atlantic Ocean and over to South America, and from Africa to the Indian Ocean and Australia, have long been predicted via models.

While measurements of a number of these transports have been captured at least spatially and temporally, it was not until the advent of spaceborne active remote sensing by lidar that characterisation of these transports in the vertical became possible. Even with spaceborne sensors, detection of aerosol plumes can only be accomplished if the aerosol concentration is sufficient to meet minimum detection thresholds. An aerosol layer that is dense and easily detectable near the source region eventually spreads and disperses beyond the minimum detectable limit for the sensor.

Hence, discerning information on aerosol and trace gas transports has been heavily reliant upon modelled information which is itself suspect in such a data-limited part of the world as the African continent and remote Indian Ocean. The CATS lidar on the ISS had detection sensitivity sufficient to identify diffuse elevated aerosol plumes originating on the African subcontinent and transported towards Australia. As a result, CATS has provided an opportunity to test and validate long-held assumptions regarding the nature of aerosol and trace gas transports in these remote regions of the world.

Authors' contributions

M.J.M.: Conceptualisation, writing (initial draft and revisions). R.J.S.: Conceptualisation, writing (initial draft and revisions). J.E.Y.: Conceptualisation, writing (initial draft). S.J.P.: Conceptualisation, writing (revisions). P.A.S.: Data analysis.

Data availability

The CATS data used in this paper are archived in NASA's Atmospheric Science Data Center (ASDC) Distributed Active Archive Center (DAAC), and are accessible via the CATS website (https://cats.gsfc.nasa.gov).

References

1. Carlson TN, Prospero JM. The large scale movement of Saharan air outbreaks over the northern equatorial Atlantic. J Appl Meteorol. 1972;11:283–297. https://doi.org/10.1175/1520-0450(1972)011<0283:tlsmos>2.0.co;2
2. Talbot RW, Harriss RC, Browell EV, Gregory GL, Sebacher DI, Beck SM. Distribution and geochemistry of aerosols in the tropical north Atlantic troposphere: Relationship to Saharan dust. J Geophys Res. 1986;91:5173–5182. https://doi.org/10.1029/JD091iD04p05173
3. Swap R, Ulanski S, Cobbett M, Garstang M. Temporal and spatial characteristics of Saharan dust outbreaks. J Geophys Res. 1996;101:4205–4220. https://doi.org/10.1029/95JD03236
4. Husar RB, Prospero JM, Stowe LL. Characterization of tropospheric aerosols over the oceans with the NOAA advanced very high resolution radiometer optical thickness operational product. J Geophys Res. 1997;102:16889–16909. https://doi.org/10.1029/96JD04009
5. Prospero JM, Glaccum RA, Nees RT. Atmospheric transport of soil dust from Africa to South America. Nature. 1981;289:570–572. https://doi.org/10.1038/289570a0
6. Talbot RW, Andreae MO, Berresheim H, Artaxo P, Garstang M, Harriss RC, et al. Aerosol chemistry during the wet season in central Amazonia: The influence of long-range transport. J Geophys Res. 1990;95:16955–16969. https://doi.org/10.1029/JD095iD10p16955
7. Swap R, Garstang M, Greco S, Talbot R, Kållberg P. Saharan dust in the Amazon Basin. Tellus B. 1992;44:133–149. https://doi.org/10.3402/tellusb.v44i2.15434
8. Kaufman YJ, Koren I, Remer LA, Tanre D, Ginoux P, Fan S. Dust transport and deposition observed from the Terra-Moderate Resolution Imaging Spectroradiometer (MODIS) spacecraft over the Atlantic Ocean. J Geophys Res. 2005;110, D10S12, 16 pages. https://doi.org/10.1029/2003JD004436
9. Koren I, Kaufman YJ, Washington R, Todd MC, Rudich Y, Martins JV, et al. The Bodélé depression: A single spot in the Sahara that provides most of the mineral dust to the Amazon forest. Environ Res Lett. 2006;1, Art. #014005, 5 pages. https://doi.org/10.1088/1748-9326/1/1/014005
10. Yu H, Chin M, Yuan T, Bian H, Remer LA, Prospero JM, et al. The fertilizing role of African dust in the Amazon rainforest: A first multiyear assessment based on data from Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations. Geophys Res Lett. 2015;42:1984–1991. https://doi.org/10.1002/2015GL063040

11. Parkin DW, Phillips DR, Sullivan RAL, Johnson LR. Airborne dust collections down the Atlantic. Q J R Meteorol Soc. 1972;98:798–808. https://doi.org/10.1002/qj.49709841807

12. Fishman J, Fakhruzzaman K, Cros B, Mganga D. Identification of widespread pollution in the southern hemisphere deduced from satellite analyses. Science. 1991;252:1693–1696. https://doi.org/10.1126/science.252.5013.1693

13. Garstang M, Tyson PD, Swap R, Edwards M, Kållberg P, Lindesay JA. Horizontal and vertical transport of air over southern Africa. J Geophys Res. 1996;101:23721–23736. https://doi.org/10.1029/95JD00844

14. Tyson PD, Garstang M, Swap R, Kållberg P, Edwards M. An air transport climatology for subtropical southern Africa. Int J Climatol. 1996;16:265–291. https://doi.org/10.1002/(sici)1097-0088(199603)16:3<265::aid-joc8>3.0.co;2-m

15. Anderson BE, Grant WB, Gregory GL, Browell EV, Collins JE Jr, Sachse GW, et al. Aerosols from biomass burning over the tropical South Atlantic region: Distributions and impacts. J Geophys Res. 1996;101:24117–24137. https://doi.org/10.1029/96JD00717

16. Browell EV, Fenn MA, Butler CF, Grant WB, Clayton MB, Fishman J, et al. Ozone and aerosol distributions and air mass characteristics over the South Atlantic Basin during the burning season. J Geophys Res. 1996;101:24043–24068. https://doi.org/10.1029/95JD02536

17. Swap R, Garstang M, Macko SA, Tyson PD, Maenhaut W, Artaxo P, et al. The long-range transport of southern African aerosols to the tropical South Atlantic. J Geophys Res. 1996;101:23777–23791. https://doi.org/10.1029/95JD01049

18. Stohl A, Eckhardt S, Forster C, James P, Spichtinger N. On the pathways and timescales of intercontinental air pollution transport. J Geophys Res. 2002;107:4684. https://doi.org/10.1029/2001JD001396

19. Tyson PD, Garstang M, Swap R. Large-scale recirculation of air over southern Africa. J Appl Meteorol. 1996;35:2218–2236.

20. Tyson PD, D'Abreton PA. Transport and recirculation of aerosols off southern Africa: Macroscale plume structure. Atmos Environ. 1998;32:1511–1524. https://doi.org/10.1016/S1352-2310(97)00392-0

21. Piketh SJ, Swap RJ, Anderson CA, Freiman MT, Zunckel M, Held G. The Ben Macdhui high altitude trace gas and aerosol transport experiment. S Afr J Sci. 1999;95:35–43.

22. Rayner PJ, Law RM. A comparison of modelled responses to prescribed CO_2 sources. CSIRO technical paper no. 36. Canberra: Division of Atmospheric Research, CSIRO; 1995. p. 82.

23. Herman JR, Bhartia PK, Torres O, Hsu C, Seftor C, Celarier E. Global distribution of UV-absorbing aerosols from Nimbus 7/TOMS data. J Geophys Res. 1997;102:16911–16922. https://doi.org/10.1029/96JD03680

24. Sturman AS, Tyson PD, D'Abreton PC. Transport of air from Africa and Australia to New Zealand. J R Soc N Z. 1997;27:485–498.

25. Piketh SJ, Tyson PD, Steffan W. Aeolian transport from southern Africa and iron fertilisation of marine biota in the south Indian Ocean. S Afr J Sci. 2000;96:244–246.

26. Rosen J, Young S, Laby J, Kjome N, Gras J. Springtime aerosol layers in the free troposphere over Australia: Mildura Aerosol Tropospheric Experiment (MATE 98). J Geophys Res. 2000;105:17833–17842. https://doi.org/10.1029/2000jd900208

27. Zunckel M, Robertson L, Tyson PD, Rodhe H. Modelled transport and deposition of sulphur over southern Africa. Atmos Environ. 2000;34:2797–2808. https://doi.org/10.1016/S1352-2310(99)00495-1

28. Young SA, Langenfelds RL, Rosen JM. Observations over Australia of plumes from distant biomass-burning sources using remote sensing and airborne techniques. In: Proceedings of IGARSS 2001: IEEE 2001 International Geoscience and Remote Sensing Symposium: Scanning the present and resolving the future; 2001 July 9–13; Sydney, Australia. Piscataway, NJ: Institute of Electrical and Electronics Engineers; 2001. p. 88–90. https://doi.org/10.1109/igarss.2001.976065

29. Piketh SJ, Swap RJ, Maenhaut W, Annegarn HJ, Formenti P. Chemical evidence of long-range atmospheric transport over southern Africa. J Geophys Res. 2003;107:4817. https://doi.org/10.1029/2002jd002056

30. Wenig M, Spichtinger N, Stohl A, Held G, Beirle S, Wagner T, et al. Intercontinental transport of nitrogen oxide pollution plumes. Atmos Chem Phys. 2003;3:387–393. https://doi.org/10.5194/acp-3-387-2003

31. Swap RJ, Annegarn HJ, Suttles JT, King MD, Platnick S, Privette JL, et al. Africa burning: A thematic analysis of the Southern African Regional Science Initiative (SAFARI 2000). J Geophys Res. 2003;108(D13), Art. #8465, 15 pages. https://doi.org/10.1029/2003JD003747

32. Pak BC, Langenfelds RL, Young SA, Francey RJ, Meyer CP, Kivlighon LM, et al. Measurements of biomass burning influences in the troposphere over southeast Australia during the SAFARI 2000 dry season campaign. J Geophys Res. 2003;108(D13), Art. #8480, 10 pages. https://doi.org/10.1029/2002JD002343

33. Duflot V, Royer P, Chazette P, Baray J-L, Courcoux Y, Delmas R. Marine and biomass burning aerosols in the southern Indian Ocean: Retrieval of aerosol optical properties from shipborne lidar and sun photometer measurements. J Geophys Res. 2011;116, D18208, 15 pages. https://doi.org/10.1029/2011JD015839

34. Lindesay JA, Andreae MO, Goldammer JG, Harris G, Annegarn HJ, Garstang M, et al. International Geosphere Biosphere Programme/International Global Atmospheric Chemistry SAFARI 92 field experiment: Background and overview. J Geophys Res. 1996;101:23521–23530. https://doi.org/10.1029/96jd01512

35. Remer LA, Kleidman RG, Levy RC, Kaufman YJ, Tanre D, Mattoo S, et al. Global aerosol climatology from the MODIS satellite sensor. J Geophys Res. 2008;113, D14S07, 18 pages. https://doi.org/10.1029/2007JD009661

36. Levy RC, Mattoo S, Munchak LA, Remer LA, Sayer AM, Patadia F, et al. The Collection 6 MODIS aerosol products over land and ocean. Atmos Meas Tech. 2013;6:2989–3034. https://doi.org/10.5194/amt-6-2989-2013

37. Sayer AM, Hsu NC, Lee J, Kim WV, Dutcher ST. Validation, stability, and consistency of MODIS Collection 6.1 and VIIRS version 1 deep blue aerosol data over land. J Geophys Res. 2019;124:4658–4688. https://doi.org/10.1029/2018JD029598

38. Winker DM, Hunt WH, McGill MJ. Initial performance assessment of the CALIOP. Geophys Res Lett. 2007;34, L19803, 5 pages. https://doi.org/10.1029/2007GL030135

39. Edwards DP, Emmons LK, Gille JC, Chu A, Attie J-L, Giglio L, et al. Satellite-observed pollution from southern hemisphere biomass burning. J Geophys Res. 2006;111, D14312, 17 pages. https://doi.org/10.1029/2005JD006655

40. Tlhalerwa K, Freiman MT, Piketh SJ. Aerosol deposition off the Southern African West Coast by berg winds. S Afr Geogr J. 2005;87:152–161. http://doi.org/10.1080/03736245.2005.9713838

41. Adebiyi AA, Zuidema P. The role of the southern African easterly jet in modifying the southeast Atlantic aerosol and cloud environments. Q J R Meteorol Soc. 2016;142:1574–1589. http://doi.org/10.1002/qj.2765

42. Kent GS, Trepte CR, Skeens KM, Winker DM. LITE and SAGE II measurements of aerosols in the southern hemisphere upper troposphere. J Geophys Res. 1998;103:19111–19127. https://doi.org/10.1029/98JD00364

43. McGill MJ, Yorks JE, Scott VS, Kupchock AW, Selmer PA. The Cloud-Aerosol Transport System (CATS): A technology demonstration on the International Space Station. In: Proceedings of SPIE 9612: Lidar Remote Sensing for Environmental Monitoring XV. 2015, 96120A. https://doi.org/10.1117/12.2190841

44. Yorks JE, McGill MJ, Palm SP, Hlavka DL, Selmer PA, Nowottnick EP, et al. An overview of the CATS level 1 processing algorithms and data products. Geophys Res Lett. 2016;43:4632–4639. https://doi.org/10.1002/2016GL068006

45. Pauly RM, Yorks JE, Hlavka DL, McGill MJ, Amiridis V, Palm SP, et al. Cloud-Aerosol Transport System (CATS) 1064 nm calibration and validation. Atmos Meas Tech. 2019;12:6241–6258. https://doi.org/10.5194/amt-12-6241-2019

46. Omar AH, Winker DM, Kittaka C, Vaughan MA, Liu Z, Hu Y, et al. The CALIPSO automated aerosol classification and lidar ratio selection algorithm. J Atmos Ocean Tech. 2009;26:1994–2014. https://doi.org/10.1175/2009JTECHA1231.1

47. Groß S, Esselborn M, Weinzierl B, Wirth M, Fix A, Petzold A. Aerosol classification by airborne high spectral resolution lidar observations. Atmos Chem Phys. 2013;13:2487–2505. https://doi.org/10.5194/acp-13-2487-2013

48. Stein AF, Draxler RR, Rolph GD, Stunder BJB, Cohen MD, Ngan F. NOAA's Hysplit atmospheric transport and dispersion modeling system. Bull Am Meteorol Soc. 2015;96:2059–2077. https://doi.org/10.1175/BAMS-D-14-00110.1

49. D'Abreton PC, Tyson P. Divergent and non-divergent water vapor transport over southern Africa during wet and dry conditions. Meteorol Atmos Phys. 1995;55:47–59. https://doi.org/10.1007/BF01029601

50. Thompson AM, Balashov NV, Witte JC, Coetzee JGR, Thouret V, Posny F. Tropospheric ozone increases over the southern Africa region: Bellwether for rapid growth in southern hemisphere pollution? Atmos Chem Phys. 2014;14:9855–9869. https://doi.org/10.5194/acp-14-9855-2014

51. Piketh SJ, Annegarn HJ, Tyson PD. Lower tropospheric aerosol loadings over South Africa: The relative contribution of aeolian dust, industrial emissions, and biomass burning. J Geophys Res. 1999;104:1597–1607. https://doi.org/10.1029/1998JD100014

52. Moody JL, Pszenny AAP, Gaudry A, Keene WC, Galloway JN, Polian G. Precipitation composition and its variability in the southern Indian Ocean: Amsterdam Island, 1980-1987. J Geophys Res. 1991;96:20769–20786. https://doi.org/10.1029/91JD01921

AUTHORS:
Omosalewa Odebiri[1]
Onisimo Mutanga[1]
John Odindi[1]
Kabir Peerbhay[1,2]
Steven Dovey[2]
Riyad Ismail[1,2]

AFFILIATIONS:
[1]Department of Geography, School of Agriculture, Earth and Environmental Science, University of KwaZulu-Natal, Durban, South Africa
[2]Sappi Forests, Howick, South Africa

CORRESPONDENCE TO:
Omosalewa Odebiri

EMAIL:
odebsconstant@gmail.com

DATES:
Received: 03 May 2019
Revised: 01 Nov. 2019
Accepted: 04 Jan. 2020
Published: 26 Mar. 2020

HOW TO CITE:
Odebiri O, Mutanga O, Odindi J, Peerbhay K, Dovey S, Ismail R. Estimating soil organic carbon stocks under commercial forestry using topo-climate variables in KwaZulu-Natal, South Africa. S Afr J Sci. 2020;116(3/4), Art. #6339, 8 pages. https://doi.org/10.17159/sajs.2020/6339

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:
commercial forests, SOC variability, Maxent, climate change mitigation

FUNDING:
DST/NRF SARChI, National Research Foundation (South Africa), Institute for Commercial Forestry Research

Estimating soil organic carbon stocks under commercial forestry using topo-climate variables in KwaZulu-Natal, South Africa

Commercial forests are expanding globally, with great potential to absorb carbon and mitigate climate change. However, whereas the role of natural forests in carbon sequestration has been widely investigated, there is a paucity in the literature on the role of commercial forests in carbon assimilation. Hence, understanding the role of commercial forests in carbon storage is essential for quantifying local, regional or global carbon balances, which is valuable for climate change mitigation. Soil carbon is known to be the largest pool within any forest landscape, and is controlled by a wide range of physical and climatic factors. However, the relationship between soil organic carbon (SOC) and topo-climatic variables controlling its distribution within commercial forests is still poorly understood. Due to the limitations encountered in traditional systems of SOC determination, particularly at large spatial extents, geospatial techniques have recently emerged as a viable alternative for mapping soil properties. Therefore, this study sought to map SOC stocks variability within the commercial forest landscape, using landscape topo-climatic variables. A total of 81 soil samples was analysed for SOC concentrations and 31 topographic and climatic variables were used as predictors to SOC variability. To reduce multicollinearity, these variables were reduced to 11 using stepwise backward elimination and the maximum entropy (Maxent) algorithm was used for regression analysis to determine the relationship between SOC and the selected topo-climatic variables. Good accuracies were obtained for both training (area under the curve = 0.906) and test (area under the curve = 0.885) data sets, and demonstrate the effectiveness of selected topo-climatic variables and the Maxent algorithm in predicting SOC stocks. This study provides a framework for monitoring the status of soil carbon in commercial forest compartments and provides a viable approach for local, national or regional carbon accounting – valuable for climate change mitigation.

Significance:

- Rainfall and temperature, as well as topographic variables (such as slope, elevation and topographic wetness index) are effective in mapping SOC distribution.
- The model developed is useful in predicting SOC occurrence and yielded an effective framework for continuous monitoring and assessment of SOC.
- The method developed in this study is cost-effective and suggests the use of other readily available climatic and topographic information for the prediction of SOC under commercial forestry in South Africa and indeed globally.
- Results from this study are important to achieve the national carbon accounting objective and are also valuable to forest managers, ecologists and relevant stakeholders in understanding the spatial distribution of SOC.

Introduction

Commercial forests represent a large carbon pool with the potential to reduce net greenhouse gas emissions.[1] The world's commercial forests increased by over 105 million ha between 1990 and 2015, while natural forests decreased by 234 million ha.[2] Federici et al.[3] notes that carbon held by commercial forests was comparable to that of natural forests (i.e. 1.08 vs. 1.44 gigatonne CO_2 per year) during this period. Recent projections have also shown that commercial forest plantations will increase by 20 to 50% by the year 2030.[4,5] Hence, commercial forests are increasingly becoming valuable environmental assets. Specifically, commercial forests could be used to reduce harvest pressure on remnant natural forests, restore degraded ecosystems, conserve natural resources and design climate change mitigation policies.[4]

Carbon in commercial forest ecosystems is stored in five different pools: above- and below-ground live tree biomass, dead wood, litter and soils.[6] These pools are dynamic and controlled by a wide range of environmental factors such as climate, topography, forest type, moisture, temperature, soil type and land use.[7] Within these pools, soil is the largest carbon reservoir and constitutes 50–80% of total carbon stocks.[6] As indicated by Liu et al.[8], the global soil organic carbon (SOC) stock has been evaluated to be about 1500 Pg carbon in the upper 100-cm soil layer, which is approximately double the measure of carbon in the atmosphere and thrice the amount stored in terrestrial vegetation.[9] However, whereas the above-ground biomass carbon pools have been widely investigated, the spatial variability of SOC within forest landscapes is still poorly understood.[10] Hence, an in-depth understanding of SOC and its variability in relation to topographical and environmental factors is crucial in quantifying regional and global carbon balances and in examining the responses and feedbacks of the terrestrial ecosystem to climate change.[8] Furthermore, knowledge of SOC variability is useful for developing suitable management strategies to improve carbon assimilation[11] and to achieve total annual national and global carbon accounting objectives as well as Intergovernmental Panel on Climate Change and Kyoto Protocol objectives.

Environmental variables are critical to the spatial distribution of SOC within a forest landscape.[12] Some of these environmental factors occur naturally while others are human-induced.[13] Variables such as topography (e.g. elevation, slope and aspect) and climate (e.g. temperature and rainfall) significantly influence SOC distribution within a forest landscape.[14] Studies have also shown that climate and topography exert a strong influence on the amount of vegetation density within a forest landscape, which in turn determines SOC distribution.[15,16] However, Chaplot et al.[11] and Liu et al.[8] note that large-scale investigations on the connection between SOC distribution and the influence of multiple topo-climatic variables at a regional and global level are not sufficient. Explicitly significant relationships and models that link topo-climatic variables with SOC processes at large spatial extents are still very necessary.[17]

Traditionally, SOC has been quantified through, among others, field surveys and wet soil analysis.[18] Although these strategies are highly accurate, they are difficult to conduct over large areas, expensive, time-consuming and labour-intensive, and may lead to the generation of toxic waste such as chromate oxidation, which requires careful and proper disposal.[19] **G**eospatial techniques, on the other hand, offer more practical and economical means of predicting and quantifying soil parameters at local, regional and even global scales.[8,20]

Recently, the adoption of spatial techniques in SOC studies has attracted significant attention[8,14,21,22], particularly because spatially continuous topographic metrics that affect SOC distribution (e.g. slope, aspect, elevation, curvatures, catchment area) can now be easily generated from digital elevation models (DEMs) and satellite imagery.[14] Climatic variables such as rainfall and temperature can also be derived from WorldClim database (global climate layers) at a spatial resolution of about 1 km^2.[23] The use of DEMs derived from satellite missions such as Light Detection and Ranging (LiDAR) and Shuttle Radar Topography Mission (SRTM) for terrain analysis has become particularly popular due to their relatively high resolutions.[24] Hence, the development of models with SRTM and LiDAR-derived topographic metrics in concert with selected bioclimatic variables could benefit studies characterised by limited observations and can be used to produce continuous SOC distribution.[21] Additionally, remotely sensed derived models (e.g. DEMs, Landsat and Sentinel sensors) are useful in determining regional and global land conditions, impacts and efficacy of land management and better at characterising forest landscapes.[14] Therefore, the objective of this study was to explore the use of remotely sensed SRTM DEM derived topographic metrics and other bioclimatic variables in the prediction of SOC stocks under commercial forestry in KwaZulu-Natal, South Africa.

Materials and methods

Study site

Commercial forest plantations in South Africa cover approximately 1.4 million hectares, representing approximately 1.1% of the country's landmass.[6,25] South African commercial forest plantations are mostly found in the eastern part of the country, stretching from the Western Cape through the Eastern Cape, to the KwaZulu-Natal, Mpumalanga and Limpopo Provinces.[6,25,26] South Africa's commercial forest plantations are of great economic and environmental value. Commercial forestry contributes about 1.27% to South Africa's gross domestic product, provides direct and indirect employment to about half a million people, supplies South African forest and fibre needs of approximately 17 million air dry tons annually, and sequestrates about 4.1 million tons of CO_2 per year.[6,27] The country's forest plantations are predominantly pine (49.8%), eucalyptus (42.7%) and wattle (7.1%) species.[25]

This study was conducted across commercial forest plantations in KwaZulu-Natal Province – a region located along the east coast of South Africa (Figure 1). The Province is an important agricultural and forestry production area and is characterised by different forest species. The dominant commercial forest species in the region are *Eucalyptus* (hardwood) and *Pinus* (softwood) species, which cover approximately 371 034 and 115 922 ha, respectively.[25] The area experiences sub-tropical climatic conditions, with the main rainfall from October all through to March. The mean annual temperature of the study region is approximately 21.7 °C while the mean annual rainfall ranges from 700 mm to 1500 mm; thus conditions are favourable for commercial forestry farming.[20,28] The topography of the area is generally characterised by undulating plains and the altitude rises from 800 m to 1400 m above mean sea level. The zone is underlain by geological formations including granite, arenite, basalt, tillite, shale, sandstones and mudstones, which result in clay to pure sandy soils.[19]

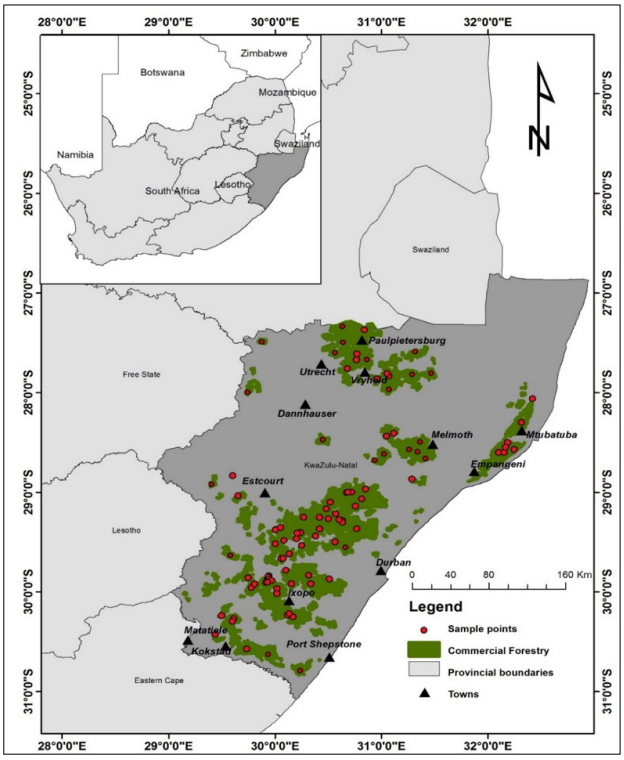


Figure 1: Commercial plantations located within KwaZulu-Natal Province, South Africa.

Field data collection and analysis

SOC analysis and quantification is an important part of forest resource management.[6] The procedure of soil sample collection in the field should be typical of the area being sampled because the utilisation of the obtained results from the laboratory analysis relies on the sampling accuracy.[29,30] Commercial forest sites are categorised into broad regions of site quality (in reference to productivity), i.e. poor, medium and good [26], hence soil samples must be taken to represent each identified site quality in the field. In this study, field data were collected during the rainy summer season (January and March 2013) – a period characterised by high above-ground biomass productivity. A stratified random sampling method was adopted, in which the area was divided into three different strata (i.e. low, medium and high productivity) based on prior knowledge of the commercial forest stands. The purpose of using a stratified random sample is to get a more representative sample of all the regions of commercial forest site qualities. A Hawth's Analysis tool in ArcGIS version 10.4.1 was used to generate random sample plots on a commercial forest cover map across all site qualities. These sample plots were allotted to the three various strata of homogeneous vegetation according to the stratified random sampling strategy and then uploaded into a handheld GPS (TRIMBLE GEO-7X) which was used to navigate to the field sites. Once the sample plots were located on the sites, soil was dug at each sample points to a depth of 30 cm using a soil auger, which is the recommended depth in spatial SOC inventories.[10,29,31] In total, 81 accessible soil samples were collected and transported to the laboratory where they were processed and analysed. SOC concentration was determined in the laboratory using the Walkley–Black32 dichromate oxidation method.

Environmental variables

Topographic metrics

Previous studies have identified topographic variables as critical drivers to SOC distribution.[14,21] According to Li et al.[14], spatial topographic variables

are categorised into three main groups: local, non-local and combined topographical variables. Local topographical variables examine the surface geometry at a point on the land surface such as slope, elevation and curvatures while non-local attributes portray relative locations of selected points, such as relief, catchment area, openness and flow accumulation. Combined attributes are an integration of both the local and non-local topographic variables such as slope length factor, topographic wetness index (TWI) and stream power index.[14] In this study, 29 different topographical variables (Table 1) that cut across the three classes (i.e. local, no-local and combined) were selected. These variables were derived from a 30-m resolution DEM created from SRTM data in SAGA GIS (2.3.2) and ArcGIS 10.4 software.

Table 1: Topo-climatic variables used in predicting the occurrence of soil organic carbon

S/n	Variable	Description	Unit
1	Topographic wetness index (TWI)	Steady state wetness index	–
2	Elevation (DEM)	Ground height	m
3	Slope	Steepness of the ground	radian
4	Aspect	Slope direction	radian
5	Mass balance index (MBI)	Mass balance index	m
6	Normalised height	Relative height and slope position	m
7	Standardised height	Relative height and slope position	m
8	Direct insolation	Potential incoming insolation	kw/m^2
9	Diffuse insolation	Solar radiation	kw/m^2
10	Visible sky	Lightening and sky view factor	kw/m^2
11	Sky view factor	Visibility	kw/m^2
12	Negative openness	Drainage features, soil water content	–
13	Positive openness	Drainage features, soil water content	–
14	LS Factor	Slope length factor effect on erosion	–
15	Catchment area	Run-off velocity and volume	m^2
16	Catchment height	Elevation above ground	m
17	Terrain ruggedness index (TRI)	Quantifies topographic heterogeneity	–
18	Plan curvature	Horizontal (contour) curvature	degrees/m
19	Profile curvature	Vertical rate of change of slope	degrees/m
20	Longitudinal curvature	Morphometric features	degrees/m
21	Cross-sectional curvature	Morphometric features	degrees/m
22	Minimum curvature	Lowest curvature	degrees/m
23	Maximum curvature	Highest curvature	degrees/m
24	General curvature	Horizontal and vertical curvature	degrees/m
25	Valley depth	Relative height	m
26	Terrain surface texture (TST)	Terrain texture of the surface	–
27	Topographic position index (TPI)	Location higher or lower than the average of the surroundings	–
28	Wind effect	Effect of wind on the surface	m/s
29	Convergence index	Shows structure of relief as a set of convergent (channels) and divergent (ridges) areas	m
30	Temperature (mean annual)	Temperature	°C
31	Rainfall (mean annual)	Rainfall	mm

Bioclimatic data

Bioclimatic data sets including rainfall and temperature are critical determinants of SOC.[13,23] In this study, mean temperature and rainfall bioclimatic variables were used in concert with topographic variables to predict SOC occurrence. The bioclimatic variables were obtained from the 1-km^2 30 arc-seconds spatial resolution of the WorldClim archives (http://www.worldclim.org/) of the global climate conditions. The WorldClim climatic data sets are long-term (30-year) mean annual measurements (containing grids including rainfall and temperature as well as other climatic layers summaries such as the wettest, driest, coldest and hottest quarters and months of the year. The derived temperature (Bio01) and rainfall (Bio12) bioclimatic layers used in this study were resampled in order to match the SRTM-derived DEM spatial resolution (i.e. from 1 km^2 to 30 m).

Statistical analysis

To determine the relationship between SOC and the derived environmental variables, SOC within the commercial forest stands was predicted using the maximum entropy (Maxent) regression algorithm software (version 3.3.3).[33]

Maximum entropy model

The maximum entropy model (Maxent) is a machine-learning algorithm proposed by Phillips et al.[33] Maxent models the probability of species presence based on environmental constraints and estimates the likelihood distribution with the maximum entropy, i.e. the most spread out distribution. Typically, Maxent produces an estimate of a probability of occurrence that ranges from 0 to 1, with 1 being the highest and 0 the lowest. It is a concise mathematical definition, and therefore can be adjusted to analyse data with efficient deterministic algorithms that are certain to give optimal probability distribution. Maxent performs efficiently, even with small sample sizes, is resistant to errors in occurrence data, and applies the presence-only data sets to estimate the suitability of habitat or the likelihood of target occurrence.[34] At the point where absence data exist for the species, a conditional model can be used to enable presence/absence modelling.[33] Maxent uses background/pseudo-absence and presence points that evaluate the environmental space for model testing. Environmental variables (continuous and categorical) and species-presence data are used to run the model and the influence of each variable can be determined from the jackknife tool in Maxent.[33]

In SOC modelling, the Maxent model begins with equal distribution and performs a number of repetitions based on the most important predictor variable until no further improvements in the spatial estimation of SOC are made.[35] The Maxent model aims to identify the maximum likelihood variability of SOC occurrences within the commercial forest stands in the study area.[36] The most uniform spread of the SOC occurrences is consequently identified by the model and selected from among many possible distributions.[37]

Predictor variables selection

One common limitation with regression is the issue of multicollinearity, which occurs when two or more predictor variables are highly correlated. Hence, it is often advisable to use the best and fewest number of predictor variables useful in building a model.[38] Stepwise regression was adopted to solve any possible multicollinearity and to select the best and fewest predictive variables for the Maxent model. Stepwise regression identifies the statistical importance of a subset of predictors through forward selection, backward elimination or a combination of both.[39] In this study, the stepwise backward elimination method was conducted using the 'stepAIC' function in the 'MASS' package of R statistic software 3.5.1 to select the best predictor variables which were then used for the Maxent final model. Backward elimination works by removing predictor variables (n=31) that are not significant to the model until the ideal number of predictor variables is obtained. The ideal number of variables was selected after the backward elimination process. These selected variables were then used to predict SOC using Maxent (version 3.3.3).

Model calibration, evaluation and validation

Settings are an integral part of building any model in order to get the best results. The SOC data (n=81) used in this study were partitioned into training (70%) and test (30%) data sets.[28] The training data (n=54) were used in the model building while the test data (n=23) were used for model validation as proposed by Phillips et al.[35] The default Maxent

settings were used to run the model. Prior to obtaining final results, visual inspection of the response curve and the difference between test and training area under the curve (AUC) values of the model were assessed to determine overfitting. The default regularisation was adopted as there was no overfitting detected. For the probabilistic model output, the 10-percentile train occurrence logistics limit command in Maxent was utilised to generate a SOC variability map. This threshold makes certain that 90% of the occurrence data have been predicted as present and that the omission error does not surpass 10%.[40]

Evaluation and validation are critical steps in any model building process. In this study, we used the random test percentage settings in Maxent, i.e. 70% of the data set was used to train and 30% to test the performance of the model. The receiver operating characteristic (ROC) curve was used to evaluate and validate the model. The ROC curve is a method that describes the performance level of probabilistic and deterministic detection of forecast systems.[36] It shows the likelihood that presence (sensitivity) is correctly ordered by the classifier (in our case predictors) as compared to the absence (specificity) of SOC. A two-dimensional space is used to generate the ROC curve by plotting the sensitivity as Y and the specificity as X. Generally, models with high accuracy have an AUC value close or equal to 1, whereas a value equal or less than 0.5 shows a model that performs no better than random.[33] Previous studies have broadly and successfully demonstrated the application of the ROC curve to quantitatively evaluate and validate the effectiveness of probability modelling.[14,21,36]

Results

Predictor variables selection

Table 2 shows the results of the stepwise backward elimination method conducted to remove redundant and highly correlated predictor variables. The procedure to select the best predictor variables was done using the 'stepAIC' function in the 'MASS' package of the R statistic software (version 3.5.1). The Akaike information criterion (AIC) acts like an examiner of the relative quality of models for a given set of data by assigning a value to the model (in this case 5.61), while the stepwise backward elimination method eliminates predictor variables that are less significant at each stage of the model until the lowest AIC value is attained (in this case -15.85). The general rule is that the lower the AIC value, the better the model. As shown in the table, 11 out of the 31 predictor variables used produced the lowest AIC value of -15.85 after the elimination procedure. Table 2 additionally shows the AIC value attached to each of the eleven selected variables, indicating the degree to which the lowest AIC value (-15.85) will increase should any of the variables be removed. Hence these selected variables are regarded as the best subset of predictors to be used in the prediction of SOC variability using the Maxent algorithm.

Table 2: Change in the Akaike information criterion (AIC) value with removal of each variable

S/N	Selected predictor variable	AIC = -15.8546
1	Longitudinal curvature	-15.7907
2	Aspect	-15.6924
3	Direct insolation	-14.6213
4	Digital elevation model (DEM)	-14.5480
5	Positive openness	-13.7504
6	Catchment area	-13.5806
7	Profile curvature	-13.3244
8	Topographic wetness index (TWI)	-10.5943
9	Rainfall (mean annual)	-7.1520
10	Slope	-5.9165
11	Temperature (mean annual)	8.6211

Table 3: Contribution (%) of each variable in predicting the occurrence of soil organic carbon

Variable	Contribution (%)
Rainfall (mean annual)	30.2
Temperature (mean annual)	22.9
Slope	16.1
Digital elevation model (DEM)	11.6
Topographic wetness index (TWI)	8.9
Direct insolation	3.1
Catchment area	2.9
Positive openness	1.8
Aspect	1.1
Profile curvature	0.8
Longitudinal curvature	0.6

Maxent model

Analysis of Maxent model omission/commission and ROC

The omission/commission and the ROC analysis is useful in examining the performance of the Maxent model. The omission/commission rate is calculated on both the training and test data sets. The general rule of Maxent is that a good model will produce omission rates that are in close proximity with the predicted omission due to the definition of the cumulative threshold. The predicted omission rate is demonstrated by a straight black line on the cumulative threshold result output of the Maxent model. As shown in Figure 2a, the omission on test data sets (turquoise line) is in close proximity with the predicted omission rate (black line) from the Maxent distribution, which signifies a good Maxent model for SOC occurrence.

Figure 2b shows the model evaluation for predicting SOC occurrence using the ROC of the randomly selected training and test data sets. The AUC ranges between 0 and 1. A good model will be closer or equal to 1 while a model with an AUC less than or equal to 0.5 signifies a poor model and was no better than random. As shown in Figure 2b, the Maxent model accomplished a high accuracy, giving AUC values of 0.906 and 0.885 for training and test data sets, respectively, and thus indicating that the developed model performed better than a random model ($p<0.005$).

Predictor variables contribution

A major strength of the Maxent algorithm is that it permits the assessment of individual predictor variables in order of their influence in the model. Table 3 shows the estimated percentage contribution of each predictor variable to the Maxent model. The increase in regularised gain is added to the contribution of the corresponding variable or subtracted from it in order to determine the estimates in each iteration of the training algorithm. The higher the percentage contribution of any of the variables, the higher its impact in predicting the SOC occurrence. The percentage contribution table generated by the Maxent model shows that 'rainfall (mean annual)' with a percentage contribution of 30.2% had the highest predictive contribution and was the most influential in predicting SOC occurrence. Also, variables such as temperature (22.9%), slope (16.1%), elevation (11.6%), TWI (8.9%) and direct insolation (3.1%) were contributors to SOC occurrence and accounted for over 60% of the total SOC occurrence. Predictor variables such as aspect, profile and longitudinal curvature had little influence on the Maxent model.

Figure 3 shows the jackknife test results; the jackknife test is used to assess each predictor variable contribution in order of importance to the Maxent model. The turquoise bars indicate the overall model accuracy when each of the predictor variables is excluded from the model, while the blue bars show the individual performance and accuracy of predictors when used in isolation. The red bars depict the overall gain of the model when all the variables are used together. As shown in Figure 3, 'rainfall (mean annual)' is the environmental predictor variable that has

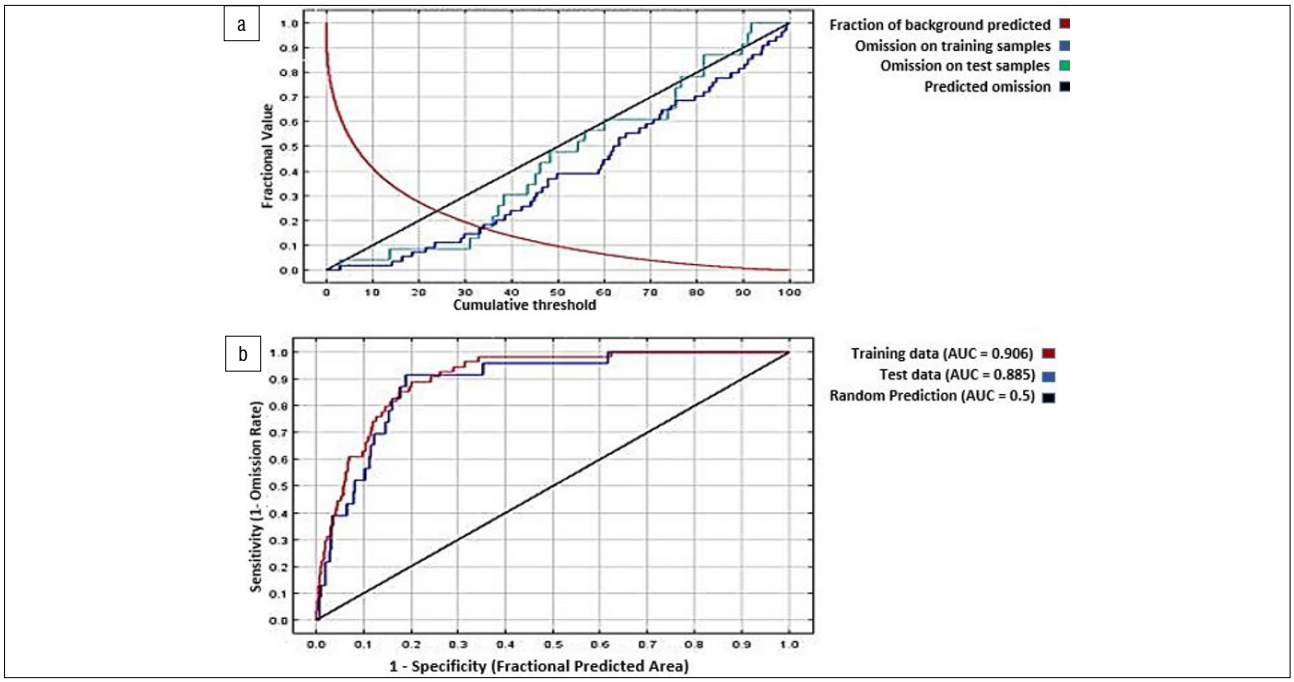


Figure 2: (a) Maxent test and training data sets omission analysis, as well as the predicted region for SOC occurrence. (b) Receiver operating characteristic (ROC) for training and test data sets using the area under the curve (AUC).

the highest gain when utilised in isolation and, as a result, appears to have the most valuable information independent of any other variable. The environmental variable that most reduced the Maxent model overall gain when omitted was 'temperature (mean annual)', which as a result appears to have the bulk of information that is absent in other variables. Variables such as slope, DEM and TWI also had an impact on the Maxent model gain as there was a significant decrease in the final model gain (red bar) when they were excluded. Other variables such as aspect, catchment area, curvatures, direct insolation and positive openness, had little or no significant contribution to the final model and hence are regarded as unimportant in predicting SOC occurrence in this study.

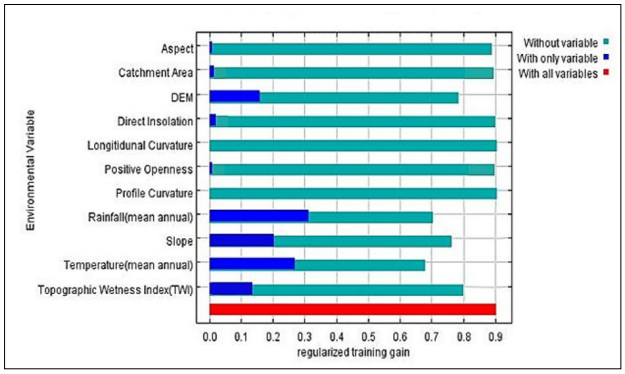


Figure 3: The jackknife of variable importance in modelling the spatial distribution of soil organic carbon.

Map generation of SOC distribution

The Maxent model generated a map that shows the spatial distribution of SOC occurrence within the study area. Figure 4 shows the likely occurrence of SOC based on the field observation points using topographic metrics and bioclimatic data as predictor variables. Areas with darker colours indicate better-predicted conditions of SOC occurrence while areas with lighter colours indicate moderate or low predicted conditions for SOC occurrence. A general assessment of the resultant SOC distribution map shows that the possible occurrence of SOC within the study region is relatively high in the southern and central plantations as compared to plantations located at the northernmost and the far eastern regions of the Province.

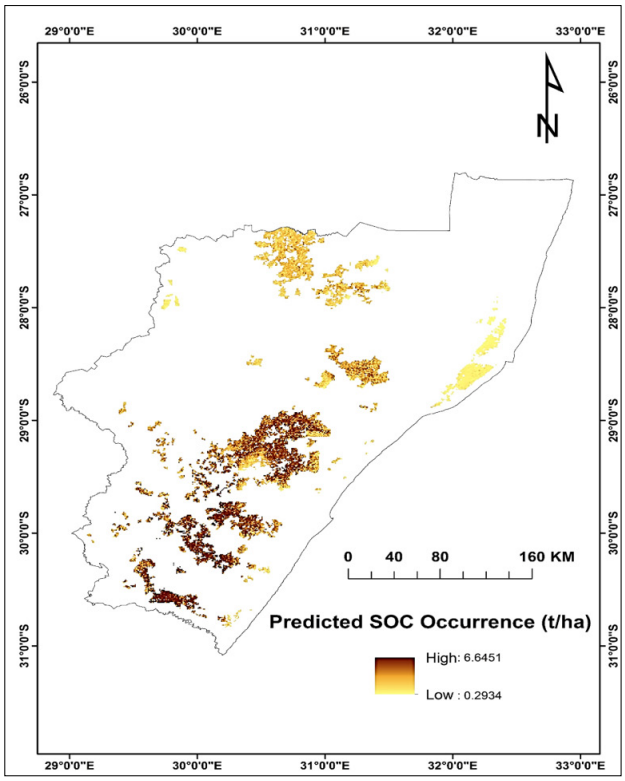


Figure 4: Map showing the spatial distribution of soil organic carbon occurrence produced by the Maxent model.

Discussion

The global expansion of commercial forestry and its potential to sequestrate carbon is increasingly becoming important in climate change mitigation.[6] Hence, there is a growing interest in commercial forest management practices to further enhance the ability to sequestrate carbon because of the carbon assimilation efficiency of commercial forests.[41] Although commercial forest soils sequestrate and store more carbon than other carbon pools, their role in carbon sequestration remains largely unexplored.[6,12,41] Therefore, we sought to estimate SOC

within commercial forestry by utilising a range of environmental variables that include topographic metrics and bioclimatic variables in a Maxent environment. Results in this study show that environmental variables such as topography and climate can be used to effectively model SOC spatial variability in a Maxent environment (AUC 0.885). The Maxent model's percentage contribution and jackknife results showed that 5 of the 11 predictor variables contributed significantly to the accuracy of the Maxent model, and hence could be used to determine SOC occurrence, distribution and variability in commercial forestry. These variables in order of importance were: rainfall, temperature, slope, elevation and TWI.

Rainfall, which contributed 30.2% to the Maxent model, was the most significant determinant of SOC occurrence in the study area. Forest stands with higher SOC occurrence are characterised by higher mean annual rainfall. This finding is consistent with that of Meier and Leuschner[42] who observed a substantial reduction (25%) in the SOC pool in forests stands that received less than 600 mm rainfall annually compared with stands that received more than 900 mm rainfall annually. Other studies, such as those of Jobbágy and Jackson[43], Bhandari and Bam[7], Chen et al.[2], Ramifehiarivo et al.[13], Hewins et al.[23] and Soucémarianadin et al.[12], also noted that mean annual rainfall within a forest landscape strongly influences the amount of SOC. The relationship between rainfall and SOC occurrence can be attributed to the influences of rainfall on soil moisture and hydrological processes such as surface run-off and groundwater infiltration.[44] Zhou et al.[45] and O'Brien et al.[46] noted that soil moisture determines SOC accumulation as it affects the vegetation density and decomposition. Furthermore, rainfall influences the presence of decomposers.[47]

The mean annual temperature also significantly influenced the prediction of SOC occurrence as it contributed 22.9% in the Maxent model. Forest stands characterised by warmer temperatures accelerate SOC mineralisation, compared with stands with lower temperatures.[48] A similar finding was reported by Wang et al.[49] who observed that soil samples incubated at different temperatures (15 °C and 75 °C) indicated higher SOC mineralisation with increasing incubation temperature. Previous investigations[13,23] have also reported temperature to be strongly linked with SOC, as it can promote plant productivity, which in turn promotes the presence of SOC. Furthermore, the influence of temperature as one of the major drivers of SOC can also be attributed to its direct relationship with rainfall, because rainfall may influence soil moisture, and by extension the surface temperature, by controlling the partitioning between sensible and latent heat fluxes.[50] In addition, the amount of rainfall within a forest landscape can also be influenced by temperature because higher temperatures signify more water vapour in the atmosphere, thus increasing the chances of heavy downpours.[51]

Slope (16.1%), elevation (11.9%) and topographic wetness index (8.9%) also showed influence on SOC. Areas with relatively steep slopes – such as the northernmost and the far eastern parts of the study area – had lower SOC occurrence than areas with relatively gentle slopes, such as the southernmost and central parts of the commercial forest areas. This finding can be attributed to higher erosion rates, which are commonly observed in areas of steeper slope as opposed to areas with less steep or gentle slopes. Erosion transports soil properties such as SOC from the upland to low-lying areas.[52] Previous studies by Li et al.[14], Fissore et al.[21], Young et al.[52] and García-Ruiz[53], among others, all noted slope as a strong determinant of SOC distribution. Elevation was also a significant influence on SOC occurrence in commercial forestry. Low likelihood SOC occurrence was predicted for areas with higher altitude (>1300 m) while high probabilities were predicted for low lying areas (<1300 m). These findings are consistent with another study[54] in which it was noted that with increasing altitude, the amount of soil and vegetation became less abundant, which was also reflected in the amount of SOC. The variability of SOC due to elevation can be attributed to the fact that low-lying areas favour vegetation growth due to optimal soil development conditions that may include erosion of nutrient-rich topsoil from higher grounds that are deposited in low-lying areas. Additionally, most low-lying areas are characterised by higher soil moisture content, nutrients and deeper depth as opposed to higher grounds characterised by extreme environmental conditions, hindering vegetation growth due to limited soil microorganisms.[55] Furthermore, the influence of elevation on microclimate could also be the reason for SOC occurrence in our study as altitude influences temperature, wind flows and soil moisture peculiar to a region, which in turn impacts the amount of vegetation and by extension SOC.

Although TWI contributed less than 10% in the Maxent model, it can still be regarded as an important driver of SOC occurrence due to its influence on the overall gains shown in the jackknife results (Figure 3). TWI determines soil moisture distribution along slopes[14], hence areas with higher TWI (soil moisture) indicate higher SOC density than areas with low TWI. Previous studies[56-58] have also reported the significance of TWI in SOC distribution by observing a strong correlation between TWI and soil organic matter. For instance, Li et al.[14] identified TWI as the most significant topographic predictor variable in SOC variability.

Results indicated relatively little impact of longitudinal and profile curvatures, catchment area, aspect, direct insolation and positive openness to SOC variability. This finding is in contrast to other investigations that demonstrated their importance to SOC formation and distribution.[14] One possible reason for the significance of these environmental factors in other studies could be due to landscape variability in relation to our study area.[21,22,39] For instance, curvatures are generally more sensitive to places of higher relief than relatively moderate landscape and have been broadly used to depict flow acceleration.[14] As a result, curvature plays a significant role in the variability patterns of SOC in regions of higher landscapes.

The present study showed that the utility of Maxent based on key topographic and bioclimatic variables provides a useful and effective methodology for predicting SOC under commercial forest landscapes. Maxent also demonstrated the ability to show the percentage contribution of each predictor variable and their influence through the analysis of the jackknife results (Table 3; Figure 3). It also automatically generated a visually appealing SOC distribution map (Figure 4) and accuracy assessments by producing the ROC curve, which is used to determine its predictive performance. Regardless of the benefits of the Maxent algorithm, it has some limitations. Maxent's logistic output relies on an assumption and not an estimation of SOC prevalence. It is also hard to compare the outputs of Maxent with other regression algorithms as it gives environmental suitability rather than predicted likelihood of SOC occurrence.

Conclusion

In this study, we investigated the impact of topographic and bioclimatic data on SOC distribution under commercial forestry using the Maxent algorithm. Results indicate that rainfall and temperature, as well as topographic variables such as slope, elevation and TWI, are effective in mapping SOC distribution. Our model was useful in predicting SOC occurrence and yielded an effective framework for continuous monitoring and assessment of SOC. The method developed in this study is cost-effective and suggests the use of other readily available climatic and topographic information for the prediction of SOC under commercial forestry in South Africa and indeed globally. Results from this study are important to achieve the national carbon accounting objective and are also valuable to forest managers, ecologists and relevant stakeholders in understanding the spatial distribution of SOC. We recommend further experiments be conducted using higher-resolution data sets to assess the performance of the Maxent algorithm in predicting SOC.

Acknowledgements

This research was supported by the South African National Research Foundation (grant no. 114898), the DST/NRF SARChI (grant no. 84157) and the Institute for Commercial Forestry Research.

Authors' contributions

O.O.: Conceived and conceptualised the study, participated in field data collection and subsequent visits to the site for ground-truthing, developed the theoretical formalism, performed the analytic calculations as well as the numerical simulations, wrote the initial manuscript draft including substantive translation. O.M.: Verified the analytical methods as well as supervised the findings of this work, contributed to results interpretation and discussion as well as participated in writing the final

manuscript, provided critical feedback and helped shape the research, analysis, and manuscript. J.O.: Participated in the formulation of the aim and objective of the study, co-supervised and contributed to the writing of every section of the manuscript, aided in interpreting the overall replication/reproducibility of results/experiments and other research outputs. K.P.: Designed field sampling procedures and techniques, coordinated the field protocol and data collection, assisted in laboratory analysis of the collected field data, participated in the overall results and discussion. S.D.: Spearheaded the laboratory analysis of collected field data, participated in the methodology section of the manuscript as well as the overall results and discussion. R.I.: Assisted in site design, coordination, and supervision, participated in subsequent site visits for ground-truthing, coordinated site mobility and equipment movement, participated in the interpretation and discussion of results.

References

1. Eggleston HS, Buendia L, Miwa K, Ngara T, Tanabe K. IPCC Guidelines for National greenhouse gas inventories. Japan: N. p.; 2006. Available from: https://www.ipcc-nggip.iges.or.jp/public/2006gl/
2. Chen X, Zhang D, Liang G, Qiu Q, Liu J, Zhou G, et al. Effects of precipitation on soil organic carbon fractions in three subtropical forests in southern China. J Plant Ecol. 2015;9(1):10–19. https://doi.org/10.1093/jpe/rtv027
3. Federici S, Tubiello FN, Salvatore M, Jacobs H, Schmidhuber J. New estimates of CO_2 forest emissions and removals: 1990–2015. Forest Ecol Manag. 2015;352:89–98. https://doi.org/10.1016/j.foreco.2015.04.022
4. Payn T, Carnus J-M, Freer-Smith P, Kimberley M, Kollert W, Liu S, et al. Changes in planted forests and future global implications. Forest Ecol Manag. 2015;352:57–67. https://doi.org/10.1016/j.foreco.2015.06.021
5. Penna I. Understanding the FAO's 'Wood Supply from Planted Forests' Projections. Ballarat, Victoria: Centre for Environmental Management, University of Ballarat; 2010.
6. Dovey SB. Current carbon stock estimation capability for South African commercial forest plantations. Report produced for the Department of Environmental Affairs. Pietermaritzburg: Institute for Commercial Forestry Research; 2014. Available from: https://www.environment.gov.za/sites/default/files/reports/carbonstocks_report2015_southafrica.pdf
7. Bhandari S, Bam S. Comparative study of soil organic carbon (SOC) under forest, cultivated and barren land: A case of Chovar village, Kathmandu. Nepal J Sci Technol. 2013;14(2):103–108. https://doi.org/10.3126/njst.v14i2.10422
8. Liu Z, Shao M, Wang Y. Effect of environmental factors on regional soil organic carbon stocks across the Loess Plateau region, China. Agric Ecosyst Environ. 2011;142(3–4):184–194. https://doi.org/10.1016/j.agee.2011.05.002
9. Smith P. How long before a change in soil organic carbon can be detected? Glob Chang Biol. 2004;10(11):1878–1883. https://doi.org/10.1111/j.1365-2486.2004.00854.x
10. Muchena R. Estimating soil carbon stocks in a dry miombo ecosystem using remote sensing. Forest Res. 2017;6(1), Art. #198, 6 pages. https://doi.org/10.4172/2168-9776.1000198
11. Chaplot V, Bouahom B, Valentin C. Soil organic carbon stocks in Laos: Spatial variations and controlling factors. Glob Chang Biol. 2010;16(4):1380–1393. https://doi.org/10.1111/j.1365-2486.2009.02013.x
12. Soucémarianadin LN, Cécillon L, Guenet B, Chenu C, Baudin F, Nicolas M, et al. Environmental factors controlling soil organic carbon stability in French forest soils. Plant Soil. 2018;426:267–286. https://doi.org/10.1007/s11104-018-3613-x
13. Ramifehiarivo N, Brossard M, Grinand C, Andriamananjara A, Razafimbelo T, Rasolohery A, et al. Mapping soil organic carbon on a national scale: Towards an improved and updated map of Madagascar. Geoderma Regional. 2017;9:29–38. https://doi.org/10.1016/j.geodrs.2016.12.002
14. Li X, McCarty GW, Karlen DL, Cambardella CA. Topographic metric predictions of soil redistribution and organic carbon in Iowa cropland fields. Catena. 2018;160:222–232. https://doi.org/10.1016/j.catena.2017.09.026
15. Bhunia GS, Shit PK, Pourghasemi HR. Soil organic carbon mapping using remote sensing techniques and multivariate regression model. Geocarto Int. 2019;34(2):215–226. https://doi.org/10.1080/10106049.2017.1381179
16. Szakács GGJ, Cerri CC, Herpin U, Bernoux M. Assessing soil carbon stocks under pastures through orbital remote sensing. Sci Agric. 2011;68(5):574–581. http://dx.doi.org/10.1590/S0103-90162011000500010
17. Fiener P, Dlugoß V, Van Oost K. Erosion-induced carbon redistribution, burial and mineralisation – Is the episodic nature of erosion processes important? Catena. 2015;133:282–292. https://doi.org/10.1016/j.catena.2015.05.027
18. Wang G, Luo Z, Han P, Chen H, Xu J. Critical carbon input to maintain current soil organic carbon stocks in global wheat systems. Sci Rep. 2016;6, Art. #19327, 8 pages. http://doi.org/10.1038/srep19327
19. Mzinyane T, Van Aardt J, Gebreslasie MT. Soil carbon estimation from *Eucalyptus grandis* using canopy spectra. S Afr J Geomatics. 2015;4(4):548–561. http://dx.doi.org/10.4314/sajg.v4i4.12
20. Dube T, Mutanga O. Evaluating the utility of the medium-spatial resolution Landsat 8 multispectral sensor in quantifying aboveground biomass in uMgeni catchment, South Africa. ISPRS J Photogramm Remote Sens. 2015;101:36–46. https://doi.org/10.1016/j.isprsjprs.2014.11.001
21. Fissore C, Dalzell B, Berhe A, Voegtle M, Evans M, Wu A. Influence of topography on soil organic carbon dynamics in a Southern California grassland. Catena. 2017;149:140–149. https://doi.org/10.1016/j.catena.2016.09.016
22. Ließ M, Schmidt J, Glaser B. Improving the spatial prediction of soil organic carbon stocks in a complex tropical mountain landscape by methodological specifications in machine learning approaches. PLoS ONE. 2016;11(4), e0153673, 22 pages. https://doi.org/10.1371/journal.pone.0153673
23. Hewins DB, Lyseng MP, Schoderbek DF, Alexander M, Willms WD, Carlyle CN, et al. Grazing and climate effects on soil organic carbon concentration and particle-size association in northern grasslands. Sci Rep. 2018;8(1), Art. #1336, 9 pages. http://doi.org/10.1038/s41598-018-19785-1
24. Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. Very high resolution interpolated climate surfaces for global land areas. Int J Climatol. 2005;25(15):1965–1978. https://doi.org/10.1002/joc.1276
25. Forestry South Africa. South African forestry and forest products industry facts [webpage on the Internet]. c2017 [cited 2019 Nov 01]. Available from: https://forestry.co.za/south-african-forestry-and-forest-products-industry-facts-1980-2016-news/
26. Smith C, Pallett R, Kunz R, Gardner R, Du Plessis M. A strategic forestry site classification for the summer rainfall region of southern Africa based on climate, geology and soils. ICFR Bull Ser. 2005;3(05).
27. South African Department of Agriculture, Forestry and Fisheries (DAFF). Report on commercial timber resources and primary roundwood processing in South Africa. Pretoria: DAFF; 2012.
28. Odindi J, Mutanga O, Rouget M, Hlanguza N. Mapping alien and indigenous vegetation in the KwaZulu-Natal Sandstone Sourveld using remotely sensed data. Bothalia. 2016;46(2), Art. #a2103, 9 pages. http://dx.doi.org/10.4102/abc.v46i2.2103
29. Ringrose S, Matheson W, Vanderpost C. Analysis of soil organic carbon and vegetation cover trends along the Botswana Kalahari Transect. J Arid Environ. 1998;38:379–396. https://doi.org/10.1006/jare.1997.0344
30. Zomer RJ, Bossio DA, Sommer R, Verchot LV. Global sequestration potential of increased organic carbon in cropland soils. Sci Rep. 2017;7(1), Art. #15554, 8 pages. https://doi.org/10.1038/s41598-017-15794-8
31. Sommer R, Bossio D. Dynamics and climate change mitigation potential of soil organic carbon sequestration. J Environ Manage. 2014;144:83–87. https://doi.org/10.1016/j.jenvman.2014.05.017
32. Walkley A, Black IA. An examination of the Degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. Soil Sci. 1934;37(1):29–38. http://dx.doi.org/10.1097/00010694-193401000-00003
33. Phillips SJ, Anderson RP, Schapire RE. Maximum entropy modeling of species geographic distributions. Ecol Modell. 2006;190(3–4):231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026
34. Graham CH, Elith J, Hijmans RJ, Guisan A, Townsend Peterson A, Loiselle BA, et al. The influence of spatial errors in species occurrence data used in distribution models. J Appl Ecol. 2008;45(1):239–247. https://doi.org/10.1111/j.1365-2664.2007.01408.x

35. Phillips SJ, Dudík M, Schapire RE, editors. A maximum entropy approach to species distribution modeling. In: Greiner R, Schuurmans D, editors. Proceedings of the Twenty-First International Conference on Machine Learning; 2004 July 4–8; Banff, Alberta, Canada. Menlo Park, CA: AAAI Press; 2004. http://doi.org/10.1145/1015330.1015412

36. Rahmati O, Pourghasemi HR, Melesse AM. Application of GIS-based data driven random forest and maximum entropy models for groundwater potential mapping: A case study at Mehran Region, Iran. Catena. 2016;137:360–372. https://doi.org/10.1016/j.catena.2015.10.010

37. Phillips SJ, Dudík M. Modeling of species distributions with Maxent: New extensions and a comprehensive evaluation. Ecography. 2008;31(2):161–175. https://doi.org/10.1111/j.0906-7590.2008.5203.x

38. Ismail R, Mutanga O. A comparison of regression tree ensembles: Predicting *Sirex noctilio* induced water stress in *Pinus patula* forests of KwaZulu-Natal, South Africa. Int J Appl Earth Obs. 2010;12:S45–S51. https://doi.org/10.1016/j.jag.2009.09.004

39. Forkuor G, Hounkpatin OK, Welp G, Thiel M. High resolution mapping of soil properties using remote sensing variables in South-Western Burkina Faso: A comparison of machine learning and multiple linear regression models. PLoS ONE. 2017;12(1), e0170478, pages. https://doi.org/10.1371/journal.pone.0170478

40. Pearson TR, Brown SL, Birdsey RA. Measurement guidelines for the sequestration of forest carbon. Gen Tech Rep NRS-18. Newtown Square, PA: US Department of Agriculture, Forest Service, Northern Research Station; 2007. https://doi.org/10.2737/NRS-GTR-18

41. Ndalowa D. Evaluation of carbon accounting models for plantation forestry in South Africa. Stellenbosch: Stellenbosch University; 2014.

42. Meier IC, Leuschner C. Variation of soil and biomass carbon pools in beech forests across a precipitation gradient. Glob Chang Biol. 2010;16(3):1035–1045. https://doi.org/10.1111/j.1365-2486.2009.02074.x

43. Jobbágy EG, Jackson RB. The vertical distribution of soil organic carbon and its relation to climate and vegetation. Ecol Appl. 2000;10(2):423–436. https://doi.org/10.1890/1051-0761(2000)010[0423:TVDOSO]2.0.CO;2

44. Aanderud ZT, Richards JH, Svejcar T, James JJ. A shift in seasonal rainfall reduces soil organic carbon storage in a cold desert. Ecosystems. 2010;13(5):673–682. http://doi.org/10.1007/s10021-010-9346-1

45. Zhou G, Guan L, Wei X, Tang X, Liu S, Liu J, et al. Factors influencing leaf litter decomposition: An intersite decomposition experiment across China. Plant Soil. 2008;311(1–2), Art. #61. https://doi.org/10.1007/s11104-008-9658-5

46. O'Brien SL, Jastrow JD, Grimley DA, Gonzalez-Meler MA. Moisture and vegetation controls on decadal-scale accrual of soil organic carbon and total nitrogen in restored grasslands. Glob Chang Biol. 2010;16(9):2573–2588. https://doi.org/10.1111/j.1365-2486.2009.02114.x

47. Young I, Ritz K. Tillage, habitat space and function of soil microbes. Soil Till Res. 2000;53(3–4):201–213. https://doi.org/10.1016/S0167-1987(99)00106-3

48. Ci E, Al-Kaisi MM, Wang L, Ding C, Xie D. Soil organic carbon mineralization as affected by cyclical temperature fluctuations in a karst region of southwestern China. Pedosphere. 2015;25(4):512–523. https://doi.org/10.1016/S1002-0160(15)30032-1

49. Wang G, Zhou Y, Xu X, Ruan H, Wang J. Temperature sensitivity of soil organic carbon mineralization along an elevation gradient in the Wuyi Mountains, China. PLoS ONE. 2013;8(1), e53914, 7 pages. https://doi.org/10.1371/journal.pone.0053914

50. Small EE, Kurc S, editors. The influence of soil moisture on the surface energy balance in semiarid environments. AGU Spring Meeting Abstracts; 2002.

51. Cong R-G, Brady M. The interdependence between rainfall and temperature: Copula analyses. Sci World J. 2012;2012, Art. #405675, 11 pages. http://dx.doi.org/10.1100/2012/405675

52. Young CJ, Liu S, Schumacher JA, Schumacher TE, Kaspar TC, McCarty GW, et al. Evaluation of a model framework to estimate soil and soil organic carbon redistribution by water and tillage using 137Cs in two US Midwest agricultural fields. Geoderma. 2014;232:437–448. https://doi.org/10.1016/j.geoderma.2014.05.019

53. García-Ruiz JM. The effects of land uses on soil erosion in Spain: A review. Catena. 2010;81(1):1–11. https://doi.org/10.1016/j.catena.2010.01.001

54. Liu L, Wang H, Dai W, Lei X, Yang X, Li X. Spatial variability of soil organic carbon in the forestlands of northeast China. J Forestry Res. 2014;25(4):867–876. https://doi.org/10.1007/s11676-014-0533-3

55. Hijmans RJ, Graham CH. The ability of climate envelope models to predict the effect of climate change on species distributions. Glob Chang Biol. 2006;12(12):2272–2281. https://doi.org/10.1111/j.1365-2486.2006.01256.x

56. Lang M, McCarty G, Oesterling R, Yeo I-Y. Topographic metrics for improved mapping of forested wetlands. Wetlands. 2013;33(1):141–155. https://doi.org/10.1007/s13157-012-0359-8

57. Lang MW, McCarty GW. LiDAR intensity for improved detection of inundation below the forest canopy. Wetlands. 2009;29(4):1166–1178.

58. Pei T, Qin C-Z, Zhu A-X, Yang L, Luo M, Li B, et al. Mapping soil organic matter using the topographic wetness index: A comparative study based on different flow-direction algorithms and kriging methods. Ecol Indic. 2010;10(3):610–619. https://doi.org/10.1016/j.ecolind.2009.10.005

AUTHORS:
Agil Katumanyane[1]
Aquillah M. Kanzi[2]
Antoinette P. Malan[1]

AFFILIATIONS:
[1]Department of Conservation Ecology and Entomology, Stellenbosch University, Stellenbosch, South Africa
[2]Department of Biochemistry, Genetics and Microbiology, Forestry and Agricultural Biotechnology Institute, University of Pretoria, Pretoria, South Africa

CORRESPONDENCE TO:
Agil Katumanyane

EMAIL:
agil.katumanyane@fabi.up.ac.za

DATES:
Received: 15 Aug. 2019
Revised: 25 Nov. 2019
Accepted: 10 Dec. 2019
Published: 26 Mar. 2020

HOW TO CITE:
Katumanyane A, Kanzi AM, Malan AP. Sciarid pests (Diptera: Sciaridae) from undercover crop production in South Africa. S Afr J Sci. 2020;116(3/4), Art. #6822, 6 pages. https://doi.org/10.17159/sajs.2020/6822

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:
Afrotropical diptera, Bradysia impatiens, invasive species, *Lycoriella ingenua, Lycoriella sativae*

FUNDING:
NemaBio (Pty) (Ltd); National Research Foundation

Sciarid pests (Diptera: Sciaridae) from undercover crop production in South Africa

Fungus gnats (sciarids) are among the most important pests in undercover crop production. They cause direct physical damage to plant roots, transfer fungal pathogens and create entry points for soil-borne plant pathogens. In 2007, *Bradysia impatiens*, an important fungus gnat pest was found in association with major tree nursery beds in the Mpumalanga and KwaZulu-Natal Provinces of South Africa and was considered invasive. In this study, eight greenhouses were surveyed in the Western Cape Province and *B. impatiens* was found to be present in all the greenhouses. Similar to the results of the previous studies, a high haplotype diversity was identified for *B. impatiens,* which may indicate multiple strain introductions into South Africa. Two other fungus gnat species, *Lycoriella sativae* and *Lycoriella ingenua* – globally important sciarid pests of mushroom cultures – were identified as new from South Africa. Through a laboratory culture, the life cycle of *B. impatiens* was observed to be approximately 21 days at 25 °C. Females laid between 100 and 250 eggs. Possible introduction sources include contaminated vegetative material and growth media, thus there maybe need to revise the importation restrictions on these commodities. The identification of two novel species of sciarid pests that have only previously been identified in the Holarctic region could further emphasise this need. However, the recent discovery of such high sciarid diversity could also be due to only a few studies having been previously done on sciarid pests in South Africa.

Significance:

- The fungus gnat species *Bradysia impatiens* is reported as present in several undercover crops in the Western Cape Province of South Africa, including cucumbers, tomatoes, chrysanthemums, mushrooms, blueberries and various herbs.
- *Lycoriella ingenua* and *Lycoriella sativae* were also found to be present in mushroom gardens – the first report of *Lycoriella* from the Afro-tropical region.
- The three fungus gnat species are some of the most important sciarid pests in undercover crop production worldwide. This study highlights the need for more studies on the distribution of these sciarids and possible invasion history.

Introduction

Dark-winged fungus gnats (Diptera: Sciaridae) are some of the major insect pests in undercover crop production and are found in greenhouses, nursery bed crops, house plants and mushroom farms, among others. They are known to cause economic losses through direct feeding on the roots, contamination of vegetative material and marketing problems.[1,2] Their secondary effects include the transmission of fungal spores[3-5], and the creation of entry points for soil-borne plant pathogens, which have in the past tended to be overlooked[1,6,7].

Fungus gnats are a problem, primarily under conditions of excessive moisture, which commonly occur during propagation[8], at which time cuttings and plugs are developing young root systems. The larvae use their prominent chewing mouth parts to feed on most plant parts, especially on young and developing plant root systems, tender roots and root hairs. Fungus gnat larval feeding results in significant physical damage to the roots and a decrease in the plant root biomass.[9] Adult flies have also been reported to transfer fungal pathogens[3,5] when they fly from one plant to another. Additionally, they cause discomfort to farm labourers,[10] which may reduce worker productivity.

The symptoms presented by the affected plants include wilting, loss of vigour, reduced vegetative development, and loss of leaves.[3] The roots generally appear to be abraded with small brown lesions.[1] The effects concerned may result in the death of the plant, especially in cases of heavy infestation, as have been reported by Springer[1] and Mansilla and Pastoriza[6], whereas the physical damage results in the weakening of the plants involved, which reduces their marketability.[9] The combined effects of pests not only reduce the resulting yield, but also to a large extent increase the production costs per hectare.[11]

Bradysia impatiens Johannsen is one of the most prominent sciarid pests in greenhouses in the world and has recently been identified in South African pine tree nursery beds as an introduced species.[12,13] In this study, *B. impatiens* is reported as a present pest in several other crops in the Western Cape Province. Other than *B. impatiens*, this study reports *Lycoriella ingenua* Dufour and *Lycoriella sativae* Johannsen as pests of mushrooms from South Africa and the Afro-tropical region for the first time. *Lycoriella ingenua* and *L. sativae* are the most important sciarid pests in mushrooms worldwide.[14,15] An overview of the current Afro-tropical fungus gnats can be found in Menzel and Smith[16].

Recently, fungus gnats were reported as concerning by many greenhouse farmers in the Western Cape Province of South Africa. This study was then conducted to determine their diversity and biology. Fungus gnat samples were collected from eight targeted greenhouses. The identification and life cycle of *B. impatiens* and its culture on artificial media, under laboratory conditions, were also investigated.

Materials and methods

Insect collection

Adult fungus gnats and larvae were collected from eight greenhouses located in the Western Cape Province of South Africa (Figure 1, Table 1), after they had been reported as concerning pests by greenhouse growers. The various crops grown in these greenhouses included cucumbers, tomatoes, different herbs, blueberries, mushrooms and chrysanthemums. The cucumber farm was the most heavily infested and was the principal target for this study.

Infested plant parts, as well as the organic growing medium were randomly collected and brought to the laboratory. The mixture of plant parts and growing media was placed in closed plastic containers that were lined with moist paper towels. The containers were kept in a growth chamber (MRC 358, Labotec) at 25 °C, and the emerging adult fungus gnats were collected and used for identification. The collections from the plant greenhouses were made between February and May 2016 while the collections from the mushroom farms where made in April 2017.

Identification of the Bradysia *species*

Adults, and in some cases first-generation larvae, were used in the molecular identification of the species by means of the *cytochrome oxidase 1 (COI)* gene. DNA was separately extracted from the intact larvae and adults using the column-based QIAamp® DNA micro-extraction kit. The following Folmer primer set was used in the polymerase chain reaction (PCR): (LCO1490) 5'-GGTCAACAAATCATAAAGATATTGG-3' and (HCO2198) 5'-TAAACTTCAGGGTGACCAAAAAATCA-3'. The primer sets are used to amplify a 658-bp fragment of the *COI* gene in a wide range of invertebrate taxa.[17] The resulting DNA sequences were edited using CLC Main Workbench (ver. 7.7.3) and blasted in on the US National Center for Biotechnology Information (NCBI)'s GenBank to determine the species identity. To determine to which clades these samples belonged, a neighbour joining tree was generated using MEGA V 6.[18] Additional sequences of fungus gnats were obtained from GenBank and added into the analysis. These sequences included *COI* sequences of *B. impatiens* (MG295935.1, KR756595.1, EU450797.1, DQ060500.1), *Bradysia nomica* (KY846435.1), *Bradysia aprica* (JX418164.1), *Bradysia longimentula* (JN378636.1), *Bradysia pallipes* (MG295853.1), *Bradysia hilaris* (MG159303.1), *Bradysia ocellaris* (MG155760.1), *Bradysia japponica* MG157878.1 and *Masakimyia pustulae* (JQ613784.1). A distance-based phylogeny was generated using the neighbour-joining method.[19] The test of phylogeny was performed using the bootstrap method and 1000 replicates. Sequence polymorphism was analysed using DnaSP v.6.[20] The p-distance was used as the substitution model with transition and transversion included. Specimens were further morphologically verified by Hans-Georg Rudzinski (Entomo-graphisches Studio, Schwanewede, Germany) and Kai Heller (independent researcher, Quickborn, Germany).

Bradysia impatiens *rearing on artificial media*

The fungus gnats that were collected from the cucumber greenhouse were used to establish a culture in the laboratory. The method used to raise the fungus gnat culture was a combination of modified protocols employed by Cloyd and Dickinson[21] and Lee et al.[22] Greenhouse media, which consisted of partially composted, 3-mm pine wood chips (sawdust), were sterilised at 40 °C overnight, and then allowed to cool to room temperature. Sterilising the substrates at this temperature ensured the death of contaminating micro- and macro-organisms, but allowed survival of the fungal spores. The substrate was then mixed with soy meal and cornmeal, in a ratio of 3:1:1, respectively. The mixture was used as the oviposition and growth medium throughout the experiments.

Glass Petri dishes (100 × 20 mm) were lined with moist filter paper (Whatman No. 1, 90 mm), filled with moist growth media, then covered and left for 24–48 h in the growth chamber, to allow for fungal colonisation. The Petri dishes were then placed in a Perspex box, along with other Petri dishes containing adult fungus gnats. The new Petri dishes were kept moist, to attract the female fungus gnats for egg laying. After 24 h, the Petri dishes (now containing eggs), were removed, covered and transferred to a growth chamber, where they were kept at a temperature of 25 °C. The Petri dish covers were fitted with thin tissue paper to prevent the accumulation of condensed moisture on top, opened frequently for aeration, and a few drops of water were added daily until the onset of pupation. Food, consisting of cornmeal and soy meal, was sprinkled on top of the Petri dishes every second or third day, depending on the needs of the larvae population involved. As soon as the adults emerged, the original Petri dishes were transferred to the Perspex box, where they were left open for the adults to emerge, mate and fly to lay eggs in new Petri dishes.

The cycle was maintained for more than 10 generations. A new container was used for each generation, to prevent contamination of the later generations, especially in respect to mites. However, controlling all types of contamination, especially mites, in the medium was difficult, as the adult fungus gnats were observed to carry mites on their bodies as they flew to new growth media. An extra effort to reduce the mite population was made through establishing cultures using fungus gnat larvae rinsed in distilled water. Even then, mites were still a constant problem in the growth medium. The Perspex box containing the fly cultures was kept at room temperature.

Table 1: List of GPS co-ordinates at which the fungus gnats were obtained, as shown in Figure 1, the host plants and the GenBank accession numbers in this study

Numbered locality	GPS	Species name	Host	Strain number	No. of sequences	GenBank accession no.
1	-33°51'21.5"S 018°59'16.5"E	*Bradysia impatiens*	Herbs; various other cuttings	–	–	NA
2	-34°09'22.0"S 019°03'21.5"E	*B. impatiens*	*Chrysanthemum* spp.; various other cuttings	B_imp21-30	10	MK860993–MK861002
3	-33°49'43.2"S 018°55'03.0"E	*B. impatiens*	Herbs	–	–	NA
4	-33°39'55.9"S 018°56'04.4"E	*B. impatiens*	Cucumbers; *Cucumis sativus* and tomatoes	B_imp31-42	12	MK861005–MK861016
5	33°48'38.6"S 18°55'23.7"E	*B. impatiens*	Blueberries; *Cyanococcus* spp.	–	–	NA
6	33°49'41.7"S 18°55'03.6"E	*B. impatiens*	Blueberries; *Cyanococcus* spp.	B_imp1-10	10	MK860972–MK860981
7	33°47'38.7"S 18°41'28.9"E	*B. impatiens*	Mushrooms; *Pleurotus ostreatus*	B_imp11-20	10	MK860982–MK860991
			Mushrooms; *P. ostreatus*	B_impAK1-2	2	MK861003–MK861004
8	33°45'38.8"S 18°48'42.9"E	*Lycoriella sativae*	Mushrooms; *P. ostreatus*	L_sat1	1	MK861017
		L. ingenua	Mushrooms; *P. ostreatus*	L_ing1-2	2	MK861018–19

Life cycle of Bradysia impatiens *under laboratory conditions*

To observe the life cycle of *B. impatiens*, an oviposition and growth medium was prepared as described above; however, finer wood chips of about 500 microns were used. The finer wood chips were obtained by means of drying out the pine sawdust at 40 °C overnight, and by blending the dry sawdust using a commercial bar blender (stainless steel; HBB250SR). The resulting powder was then sieved through a 500-μm sieve. The smaller pine wood chips permitted easy observation and the counting of the eggs laid by each female fungus gnat. The smaller cell culture dishes (35 x 10 mm CELLSTAR®) that were used for the purpose were lined with 3-cm moist filter papers (Whatman®). Freshly emerged (neonate) adult fungus gnats (1 female and 2 males) were aspirated into the growth media, where they could mate and oviposit. After the death of the females, the eggs were counted using a light microscope, and oviposition behaviour was observed. It was important to count the eggs immediately, because fungi colonisation in the media – resulting from the dead adults – would make it difficult to observe the egg-laying patterns or to count the eggs later. The Petri dishes were then fitted with fine tissue paper on the Petri dish cover, and transferred to the growth chamber to allow for the emergence and growth of the fungus gnat larvae. Individuals from each growth stage – eggs, different larval instars, pupae and adults – were randomly removed from the medium and imaged using a Zeiss Stereo Discovery.V8 microscope fitted with an Axiocam ERc 5s (n=20). The ZEISS Labscope App for iPad was then used to measure the length and width of each selected individual.

Results

Insect collection

All insects collected were from greenhouses in the Western Cape Province, South Africa (Figure 1, Table 1). The different host crops include chrysanthemums, herbs, cucumbers, tomatoes, blueberries and mushrooms.

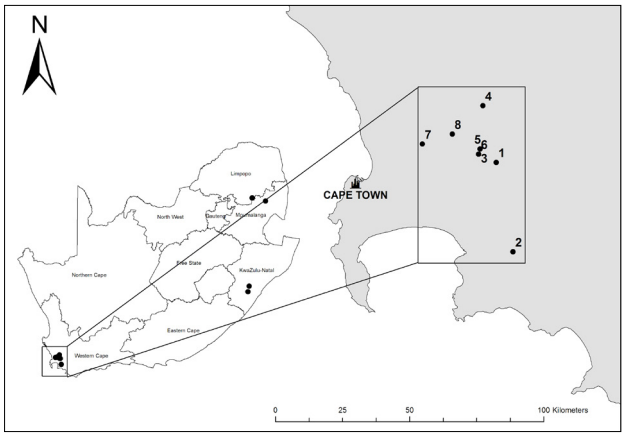


Figure 1: Localities from which fungus gnats have been reported in South Africa. Localities in the Western Cape Province are from the current study while localities in Mpumalanga and KwaZulu-Natal Provinces are from Hurley et al.[12]

Identification of Bradysia *and* Lycoriella *species*

Figure 2 shows the morphology of the three identified species. The specimens were identified morphologically using the male genitalia and the length of the antennae. Male *B. impatiens* are characterised by a comb-like row of bristles on the fore tibia and the antennal segments are more compact. *Lycoriella ingenua* is easily recognised by the conspicuous bristles on the basal lobe of male genitalia, while *Lycoriella sativae* is characterised by a fuscous, darker hypopygium and longer antennae.

For molecular identification, the resulting DNA sequences were blasted on GenBank which resulted in a 100% match with *B. impatiens*, *L. sativae* and *L. ingenua*. These sequences were submitted in the NCBI GenBank database (Table 1). The results from the neighbour-joining tree showed that the *B. impatiens* grouped together with other *Bradysia* species. It seems that there are variations in the sequences as observed in the neighbour-joining tree, which shows sub-groups within the *B. impatiens* clade, although unsupported by bootstrap support values. Analysis of sequence polymorphism revealed eight polymorphic sites (Table 2).

Table 2: Polymorphic sites identified in seven haplotypes of *cytochrome oxidase 1* gene (primer set: LCO1490 and HCO2198) of *Bradysia impatiens*. Shaded regions show haplotypes that were collapsed into Hap2 when a 350-bp (portion amplified by primers CI-J-1751 and CI-N-2191) sequence alignment was used.

Haplotype	Site							
	6	63	68	85	453	515	610	616
Hap1	C	A	A	C	C	T	A	T
Hap2	–	–	–	–	–	–	G	C
Hap3	–	–	–	T	–	–	G	C
Hap4	–	–	–	–	–	C	G	C
Hap5	–	G	–	T	–	–	G	C
Hap6	T	–	G	–	T	–	G	C
Hap7	T	–	G	–	–	–	G	C

Based on these polymorphisms, these sequences were grouped into seven haplotypes: Hap1 (B_imp1-10), Hap2 (B_imp11-23), Hap3 (B_imp24-25), Hap4 (B_imp36), Hap5 (B_imp37), Hap6 (B_imp38-44) and Hap7 (B_imp45). The haplotype (gene) diversity was 0.788 and the nucleotide diversity was 0.00325. Sequences homologous to *L. sativae* and *L. ingenua* grouped together, but separately from the *Bradysia* species (Figure 3). Haplotype analysis using a portion of sequence (350 bp) amplified by the primers CI-J-1751 and CI-N-2191 used in a previous study by Hurley et al.[23] to analyse a population of *B. difformis,* revealed four haplotypes including Hap1 (B_imp1-10), Hap2 (B_imp11-35,37,45), Hap3 (B_imp36) and Hap4 (B_imp38-44). In this data set, the haplotype diversity was 0.579 while the nucleotide diversity was 0.00258.

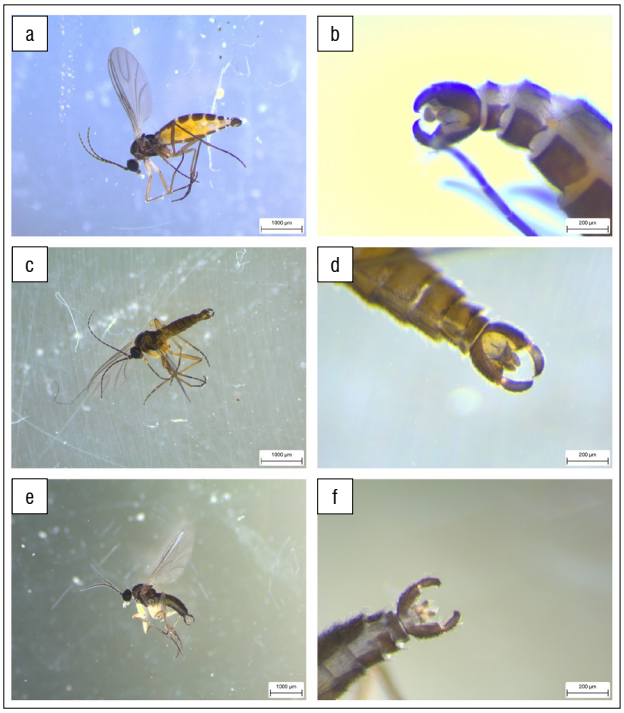


Figure 2: (a) Male and (b) male genitalia of *Lycoriella ingenua*. (c) Male and (d) male genitalia of *Lycoriella sativae*. (e) Male and (f) male genitalia of *Bradysia impatiens.*

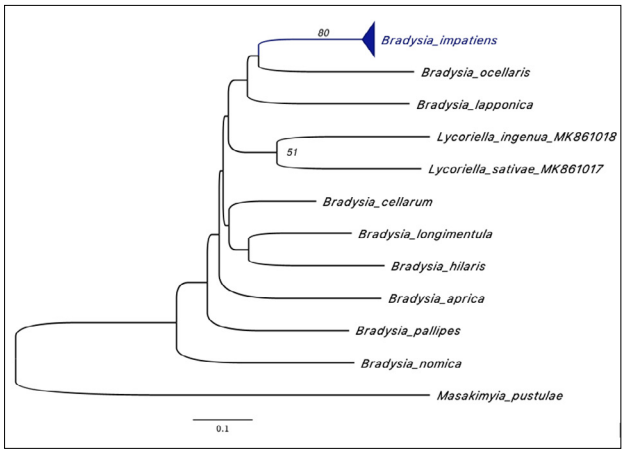


Figure 3: Neighbour-joining tree showing grouping of identified *Bradysia impatiens*, *Lycoriella ingenua* and *L. sativae* sampled and analysed in this study (Table 1). The node denoted as a blue triangle represents all the *Bradysia impatiens* sequences. Other *Bradysia* species and *Masakimyia pustulae* were included in the neighbour-joining tree for comparative purposes and to provide structure. Only bootstrap values over 50% are shown.

Rearing of Bradysia impatiens *on artificial media*

The culture medium used served as an ideal oviposition and growth medium for the fungus gnats. Up to 10 generations of fungus gnats were maintained on the same constitution of the culturing medium, without any changes being noticed in the behaviour of the fungus gnat colony. A diagram of the observed life cycle is shown in Figure 4.

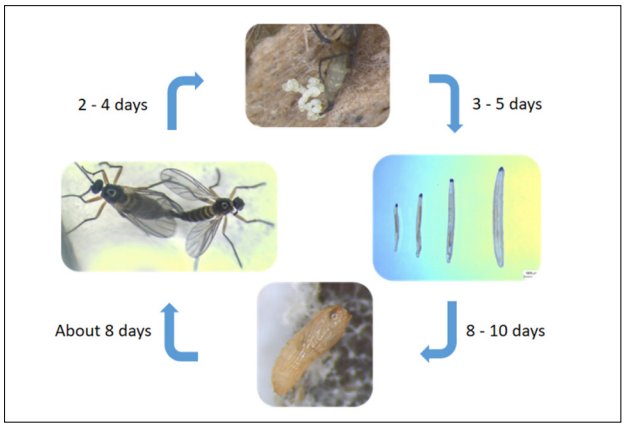


Figure 4: Observed life cycle of *Bradysia impatiens* at 25 °C.

Life cycle of Bradysia impatiens *under laboratory conditions*

The adults

Newly emerged adults had a pale neonate body with wings that barely covered the abdomen. The wings attained full length in under 30 min after emergence. The emerged adults were very active. Within a few hours after emergence, the adults had a dark-brown or dark body colour and dark wings. Mating occurred within the first few hours after emergence. The body size of the adults was 2–3 mm, with a wingspan of about 7 mm. Males were slightly smaller than females. The males hovered on top of the medium, with a bent abdomen, as they waited to mate with the emerging females. After mating, the females lived for approximately 2–4 days, after which they laid eggs and died, either immediately after or during the process of egg laying. They lowered their abdomens into crevices to lay their eggs on top of the culturing medium, or just below the moist wooden chips. If the medium particles were small enough, as was the case with the 500-μm wood chips, the females lowered their abdomens into the medium, thus completely covering the eggs.

Eggs

Egg laying was variable among females, which tended to lay eggs either in groups of variable numbers or singly, while moving on top of the medium, or at once, either in clusters or in chains, as shown in Figure 5a. The eggs were oval, with a shiny, semi-transparent light-yellowish colour, as soon as they were laid, but changed to colourless towards hatching. Before hatching, the larvae, with their prominent black heads, could be seen actively moving around and eating through the eggshell (Figure 5b). The eggs seemed to turn more yellow only under moisture stress. The eggs measured about 0.25 × 0.15 mm. The number of eggs laid by the females was variable and ranged between 100 and 250 (n=20).

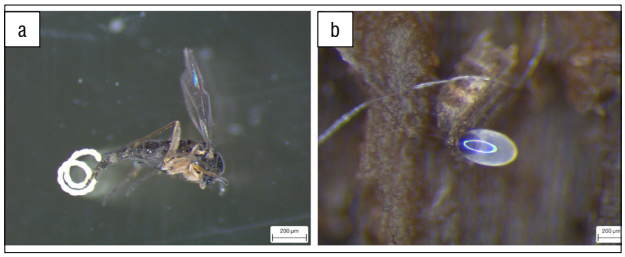


Figure 5: (a) Egg-laying pattern in the form of a chain – observed for the first time in *Bradysia impatiens*. (b) First instar larva of *B. impatiens* ready to emerge from the egg.

First instar larvae

After eating through the eggshell, the first instar became active almost immediately. Their body was transparent, while their head was shiny, black and chitinised. The food and its colour could be seen through the semi-transparent abdomen. The larvae quickly moved to the bottom of the medium, becoming active feeders. They were capable of surviving completely submerged in water, but showed a high level of susceptibility to moisture stress. The first instar larvae measured about 1.2–2.5 mm in length and 0.2–0.3 mm in width (n=20). The first instar stage lasted 2 days, after which the larvae hatched into second instar larvae.

Second instar larvae

The second instar larvae were not behaviourally different from the first instars, but differed in size: they measured about 2.6–4.5 mm in length, and 0.3–0.45 mm in width. The second instar larval stage lasted about 2 days, after which they hatched into third instar larvae.

Third instar larvae

The third instar larvae were more easily visible to the naked eye than were the earlier larval instars, and they could be seen eating through the medium. Raw food substances could also be seen inside their abdomen, thus the larvae sometimes took on the colour of the food. They measured about 4.6–6.5 mm in length, and about 0.46–0.65 mm in diameter (n=20).

Fourth instar larvae

The fourth instars were thicker than the previous instars, and contained almost no raw food substances, as they prepared to pupate. Thus, they appeared whiter than the previously described instars. They measured about 6.5–7.2 mm in length and ranged from 0.65 to 0.75 mm in diameter (n = 20). The fourth instar larvae migrated to the topmost layer of the medium, with their body continuously shortening, and thereafter pupated. This stage lasted approximately 2 days. As the larvae moved, they left behind a characteristically slimy, shiny translucent gel. All the larvae were sensitive to light, and tended to hide under the medium when they were suddenly exposed to the microscope light.

Pupa

The pupae measured about 2–3 mm in length, and 0.6–0.8 mm in diameter (n=20). The pupae cocooned within the top layers of the medium, and on approximately the third day, the adult fungus gnats emerged. The pupae were yellowish-brown in colour, but they turned dark towards emergence as adults.

Discussion

A recent study by Lee et al.[22] clarified that a few fungus gnat species, belonging to the subfamily Megalosphyinae (genus *Phytosciara* sensu lato, and part of the genus *Bradysia* sensu lato), have more frequently had their larvae associated with living plants than has any other Sciaridae group. The larvae of these fungus gnat species are capable of mining into the roots, stems and leaves of living plants. Some sciarid species are important pests of certain crops that are of global agricultural importance, especially crops under cover. *Bradysia impatiens*[24], which is a major crop pest, has a global distribution. The species has previously been identified in South Africa in association with major forestry nurseries[12,13] where the effects of these species still tend to go unnoticed and unreported by many farmers. Only a few studies have been undertaken thus far for the purpose of identification, and this study is the first on the biology of the species in South Africa. Further studies are needed on the pest status and possible management strategies of the species.

During the course of this study, *B. impatiens* was identified as present and in concerning numbers on cucumber plants in a commercial greenhouse farm in the Western Cape Province of South Africa. *Bradysia impatiens* was also identified as present in the rest of the sampled greenhouses in which tomatoes, mushrooms, blueberries, herbs and chrysanthemums were grown. Even though fungus gnats have previously been identified in South Africa in forestry nursery beds[12,13], the current study presents the first report of fungus gnats on a crop other than pine nursery beds. This study also reports *L. sativae* and *L. ingenua* as present on mushrooms, for the first time from South Africa. *Lycoriella sativae* is a well-known mushroom pest which has to date been identified to be well distributed in the Holarctic region and reported to have been distributed by humans to Central America and the sub-Antarctic islands.[24] Although *L. ingenua* has been determined to be a more important pest species in mushrooms, *L. sativae* is known to be an agrarian species and the most abundant species of Sciaridae on fields.[15,24]

The biology of *B. impatiens* is known to differ according to different environmental conditions. The duration of the life cycle of sciarids has been regarded as distinctive by a variety of researchers. Such duration has been attributed mainly to the variable environmental temperatures, with shorter life cycles at relatively high environmental temperatures.[6,25-27] The optimum temperature for their growth has been determined to be 30 °C.[2] In the current study, the duration of the life cycle of a laboratory culture of *B. impatiens* was found to be 2–3 weeks at 25 °C. On observing the egg-laying habits of the species, apart from their commonly observed habits, a new phenomenon was detected, that is, females were observed to lay their eggs in the form of a chain.

Different larval stage feeds have been described for the rearing of *Bradysia* spp. in the laboratory. Some of such feeds include the use of the fungal culture of *Pleurotus astreatus* Kumm 1871 grown on potato dextrose agar[28]; a mixture of moist coconut coir dust, commercial rabbit food and brewer's yeast; peat[8]; potato agar[6]; bacto-agar; and brewer's yeast[29]. In all the combinations mentioned, fungi make up a basic element in the feed for fungus gnats. Upon death, the adult flies were observed to act as a primary source of fungi in the media as they deteriorated. This observation suggests that, even without a fungal culture, one could easily rely on the adult fungus gnats for the introduction of fungi into the new medium. The culture medium that was used in the current study, consisting of a 3:1:1 mixture of pine sawdust, soy meal and cornmeal, gave satisfactory results for more than 10 generations. Because some fungus gnat species have been shown to prefer laying eggs on cut planes, rather than on whole stems,[22] the use of pine sawdust (blended) provided an ideal medium for egg laying.

During culturing, moisture was an important aspect in the growth of *B. impatiens,* especially in the case of the larval instars, with the first instars in the current study being observed to be the most susceptible to moisture stress. Consequently, water was added to the mixture more regularly during the larval stages. Aeration of the medium is equally important, with fungus gnats having been observed to not survive under anaerobic conditions. In the present experiment, the Petri dish cover was fitted with tissue paper to prevent the accumulation of condensed moisture on the upper lid, which otherwise was observed to be capable of creating anaerobic conditions if the condensed water created a water film on the edges of the Petri dish cover.

The fungus gnat species responsible for damage on a commercial cucumber farm was identified as *B. impatiens*. Under laboratory conditions, the life cycle of the species concerned was determined to be 2–3 weeks at 25 °C. This report is also the first of fungus gnats laying eggs in the form of a chain, as well as the first report of *B. impatiens* being a pest on any other crops than tree nursery beds in South Africa. The number of eggs laid by each single female, and their short life cycle, show that large numbers of fungus gnats can easily build up in a greenhouse if control measures are not taken.

Bradysia impatiens has been reported to be an introduced species to South Africa. A study by Hurley et al.[23] suggested that multiple strains of the species had been introduced into the country. Our results corroborate the results of this study where high haplotype diversity and comparable nucleotide diversity in the sequences analysed was observed. Multiple strain introductions could possibly have been caused by the human importation of contaminated material such as substrates and ornamental plants, among others. These findings emphasise the need for more stringent restrictions on the importation of vegetative material, and for constant monitoring of such crop pests.

From the results of the current study, as well as on the basis of the findings of previous studies, *B. impatiens* can be concluded to be a well-established pest of protected crops in South Africa. However, the current study only targeted specific greenhouses to survey fungus gnats. Such a limitation implies that more surveys are required to establish the existence of the species in other regions of South Africa, and the extent to which the species is a problem to South African growers. The question remains if these are truly introduced species or if they are South African species that have only been identified recently. If they are introduced species, it would be paramount to look into the invasion biology of these species in South Africa, as *B. impatiens* is already known to have a wide distribution in South Africa, since its first report in 2007. There is also a need to look into the management practices for fungus gnats as they are polyphagous and economic pests in undercover production. More important still, is the need for studies regarding the sustainable management of the pests concerned, because they attack crops to which the application of chemicals is relatively inappropriate. Current efforts for South Africa have included a review on the potential for biological control of fungus gnats using entomopathogenic nematodes as well as laboratory and field trials to test for the potential of South African entomopathogenic nematodes to control fungus gnats.[30,31]

Acknowledgements

We thank Hans-Georg Rudzinski (Entomo-graphisches Studio, Schwanewede, Germany) and Kai Heller (Quickborn, Germany) for assistance in the identification of fungus gnats. The financial support of NemaBio (Pty) (Ltd) and the National Research Foundation of South Africa (THRIP TP14062571871) are greatly appreciated. We also thank Sheila Storey of Nemlab for assistance during field work and Francois Bekker for drawing the map.

Authors' contributions

A.K.: Conceptualisation; methodology; data collection; sample analysis; data analysis; validation; data curation; writing – the initial draft; writing – revisions. A.M.K.: Molecular data analysis. A.M.: Conceptualisation; writing – revisions; student supervision; project leadership; project management; and funding acquisition.

Data availability

All sequences used in this study are available by searching the GenBank accession numbers provided in Table 1 on the US National Center for Biotechnology Information (NCBI) GenBank database (https://www.ncbi.nlm.nih.gov/). The sequence alignments and phylogenetic tree are available upon request.

References

1. Springer TL. Fungus gnat (Diptera: Sciaridae) feeding damage to legume seedlings. J Kans Entomol Soc. 1995;68(2):240–242. http://www.jstor.org/stable/25085588
2. Chandler D, Hri W, Prince G, Bennison J. New approaches to microbial control of insect pests in protected crops and their interaction with waste-based growing media. PC 283. Coventry: The University of Warwick; 2010.
3. Pundt L. Fungus gnats are serious pests. Yankee Grower. 1999; September/October.
4. Ludwig SW, Oetting RD. Evaluation of medium treatments for management of *Frankliniella occidentalis* (Thripidae: Thysanoptera) and *Bradysia coprophila* (Diptera: Sciaridae). Pest Manag Sci. 2001;57(12):1114–1118. https://doi.org/10.1002/ps.404
5. Scarlett K, Tesoriero R, Daniel R, Guest D. Sciarid and shore flies as aerial vectors of *Fusarium oxysporum* f. sp. *cucumerinum* in greenhouse in cucumbers. J Appl Entomol. 2014(5);138:368–377. https://doi.org/10.1111/jen.12098
6. Mansilla JP, Pastoriza MI. Estudio sobre la biología y control de *Bradysia paupera* Tuomikoski (= *Bradysia difformis* Frey, Diptera : Sciaridae) [Study on the biology and control of *Bradysia paupera* Tuomikoski (= *Bradysia difformis* Frey, (Diptera: Sciaridae)]. Bol Sanid Veg Plagas. 2001;27:411–417. Spanish.
7. Vänninen I. Control of sciarid flies with *Steinernema feltiae* in poinsettia cutting production. Int J Pest Manag. 2003;49(2):95–104. https://doi.org/10.1080/09670870210000341 39
8. Jagdale GB, Casey ML, Grewal PS, Lindquist RK. Application rate and timing, potting medium, and host plant effects on the efficacy of *Steinernema feltiae* against the fungus gnat, *Bradysia coprophila*, in floriculture. Biol Control. 2004;29(2):296–305. https://doi.org/10.1016/S1049-9644(03)00164-6
9. Cloyd RA, Zaborski ER. Fungus gnats, *Bradysia* spp. (Diptera: Sciaridae), and other arthropods in commercial bagged soilless growing media and rooted plant plugs. J Econ Entomol. 2004;97(2) 503–510. https://doi.org/10.1603/0022-0493-97.2.503
10. Schuhli GSE, Penteado SDC, Reis FW, Amorim DS. Sciarid fungus-gnats as nuisance factor in *Pinus* timber yards. Pesq flor bras. 2014;34(80):455–457. https://doi.org/10.4336/2014.pfb.34.80.732
11. Popp J, Hantos K. The impact of crop protection on agricultural production. Stud Agric Econ. 2011;113:47–66. https://doi.org/10.7896/j.1003
12. Hurley BP, Govender P, Coutinho TA, Wingfield BD, Wingfield MJ. Fungus gnats and other *Diptera* in South African forestry nurseries and their possible association with the pitch canker fungus. S Afr J Sci. 2007;103(1):43–46.
13. Hurley BP, Slippers B, Coutinho TA, Wingfield BD, Govender P, Wingfield MJ. Molecular detection of fungi carried by *Bradysia difformis* (Sciaridae: Diptera) in South African forestry nurseries. South Hemisphere For J. 2007;69:103–109. https://doi.org/10.2989/SHFJ.2007.69.2.5.291
14. Chidziya E, Mutangadura D, Jere J, Siziba L. A comparative evaluation of locally available substrates for rearing and studying biology of sciarid fly, *Lycoriella mali*. Afr J Biotechnol. 2013;1(4):57–61. http://dx.doi.org/10.15413/ajb.2013.0108
15. Menzel F, Smith JE, Colauto NB. *Bradysia difformis* Frey and *Bradysia ocellaris* (Comstock): Two additional neotropical species of black fungus gnats (Diptera: Sciaridae) of economic importance: A redescription and review. Ann Entomol Soc Am. 2003;96(4):448–457. https://doi.org/10.1603/0013-8746(2003)096[0448:BDFABO]2.0.CO;2
16. Menzel F, Smith JE. Sciaridae (black fungus gnats). In: Kirk-Spriggs AH, Sinclair BJ, editors. Manual of Afro-tropical Diptera. Pretoria: South African National Biodiversity Institute; 2017. p. 557–580.
17. Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Mol Mar Biol Biotechnol.1994;3(5):294–299.
18. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol Biol Evol. 2013;30(12):2725–2729. https://doi.org/10.1093/molbev/mst197
19. Saitou N, Nei M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol. 1987;4(4):406–425. https://doi.org/10.1093/oxfordjournals.molbev.a040454
20. Rozas J, Ferrer-Mata A, Sánchez-Delbarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, et al. DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol Biol Evol. 2017;34(12):3299–3302. https://doi.org/10.1093/molbev/msx248
21. Cloyd RA, Dickinson A. Effect of *Bacillus thuringiensis* subsp. *israelensis* and neonicotinoid insecticides on the fungus gnat *Bradysia* sp. nr. *coprophila* (Lintner) (Diptera: Sciaridae). Pest Manag Sci. 2006;62(2):171–177. https://doi.org/10.1002/ps.1143
22. Lee H, Shin SJ, Hong GT, Ahn JY, Cho HD. Biological characteristics of the ginseng stem fungus gnat (*Phytosciara procera*) and its environmental-friendly control using modified topping of ginseng peduncles. J Ginseng Res. 2010;34(1):23–29. https://doi.org/10.5142/JGR.2010.34.1.023
23. Hurley BP, Slippers B, Coutinho TA, Wingfield BD, Govender P, Wingfield MJ. Genetic diversity of *Bradysia difformis* (Sciaridae: Diptera) populations reflects movement of an invasive insect between forestry nurseries. Biol Invasions. 2009;12(4):729–733. https://doi.org/10.1007/s10530-009-9509-1
24. Mohrig W, Heller K, Hippa H, Vilkamaa P, Menzel F. Revision of black fungus gnats (Diptera: Sciaridae) of North America. Stud Dipterol. 2013;19:141–286. https://doi.org/10.11646/zootaxa.4150.4.3
25. Kennedy MK. Scientific notes: A culture method for *Bradysia impatiens* (Diptera: Sciaridae). Ann Entomol Soc Am. 1973;66(5):1163–1164. https://doi.org/10.1093/aesa/66.5.1163
26. Nielsen GR. Fungus gnats EL 50 [webpage on the Internet]. c1997 [cited 2019 Nov 25]. Available from: http://pss.uvm.edu/ppp/pubs/el50.htm
27. Villanueva-Sánchez E, Ibáñez-Bernal S, Lomelí-Flores JR, Valdez-Carrasco J. Identificación y caracterización de la mosca en el cultivo de nochebuena (Euphorbia pulcherrima) en el centro de México [Identification and characterization of the fly in the poinsettia (*Euphorbia pulcherrima*) crop in central Mexico]. Acta Zool Mex. 2013;29:363–375. Spanish.
28. Kim HH, Choo HY, Kaya HK, Lee DW, Lee SM, Jeon HY. *Steinernema carpocapsae* (Rhabditida: Steinernematidae) as a biological control agent against the fungus gnat *Bradysia agrestis* (Diptera: Sciaridae) in propagation houses. Bio-control Sci Technol. 2004;14(2):171–183. https://doi.org/10.1080/09583150310001655693
29. Katumanyane A, Ferreira T, Malan AP. *Bradysia* species (Diptera: Sciaridae) as pests of covered crops, with special reference to biological control using entomopathogenic nematodes. Afr Entomol 2018;26(1):1–13. https://doi.org/10.4001/003.026.0001
30. Katumanyane A, Ferreira T, Malan AP. Greenhouse application of *Steinernema yirgalemense* to control fungus gnats, *Bradysia impatiens* (Diptera: Scaridae). BioContr 2018;63(5):729–738. https://doi.org/10.1007/s10526-018-9895-3
31. Katumanyane A, Ferreira T, Malan AP. Potential use of local entomopathogenic nematodes to control *Bradysia impatiens* (Diptera: Sciaridae) under laboratory conditions. Afr Entomol. 2018;26(2):337–349. https://doi.org/10.4001/003.026.0337

AUTHORS:
Charles W. Helm[1,2]
Hayley C. Cawthra[1,3]
Xander Combrink[4]
Carina J.Z. Helm[5]
Renée Rust[6]
Willo Stear[1]
Alex van den Heever[7]

AFFILIATIONS:
[1]African Centre for Coastal Palaeoscience, Nelson Mandela University, Port Elizabeth, South Africa
[2]Peace Region Palaeontology Research Centre, Tumbler Ridge, British Columbia, Canada
[3]Geophysics and Remote Sensing Unit, Council for Geoscience, Cape Town, South Africa
[4]Department of Nature Conservation, Tshwane University of Technology, Pretoria, South Africa
[5]Department of Geography, University of British Columbia, Vancouver, British Columbia, Canada
[6]School of Geography, Archaeology and Environmental Studies, University of the Witwatersrand, Johannesburg, South Africa
[7]Tracker Academy, Magazine Hill, Graaff-Reinet, South Africa

CORRESPONDENCE TO:
Charles Helm

EMAIL:
helm.c.w@gmail.com

DATES:
Received: 19 July 2019
Revised: 08 Nov. 2019
Accepted: 08 Nov. 2019
Published: 26 Mar. 2020

HOW TO CITE:
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

ARTICLE INCLUDES:
☒ Peer review
☐ Supplementary material

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

EDITOR:
Maryna Steyn

KEYWORDS:
Pleistocene reptiles, fossil tracks, swim traces, crocodylians, lithics

FUNDING:
None

Pleistocene large reptile tracks and probable swim traces on South Africa's Cape south coast

The Cape south coast of South Africa contains a wealth of Pleistocene vertebrate trace fossil sites in aeolianites and cemented foreshore deposits. Published studies have described mammal and avian tracksites identified along this coastline. We report here on a number of Pleistocene palaeosurfaces within the Garden Route National Park that exhibit tracks of large reptiles, including probable swim traces. The tracks were probably made by more than one species, and may include a crocodylian. There are no extant reptiles in this coastal region capable of making such tracks and traces, which probably represent an indication of a previously more extensive range for the Nile crocodile and a monitor lizard. These findings demonstrate the potential for ichnology to complement the traditional body fossil record. Two Middle Stone Age stone artifacts were found embedded in one palaeosurface containing multiple reptile trackways. These discoveries have implications for the understanding of Pleistocene palaeoenvironment and palaeoclimate – in an area which is important in the study of modern human origins.

Significance:

- Large reptile Pleistocene fossil tracksites have recently been discovered on the Cape south coast of South Africa where there are no previous such records, and no reptiles of this size are currently found in the region.
- These sites include the first reported probable reptile swim traces in Africa and one tracksite also contained two Middle Stone Age artifacts.
- These discoveries have implications for Pleistocene environments and climate on the Cape south coast.

Introduction

Fossil tracks and trackways are common in Middle and Late Pleistocene coastal aeolianites and cemented foreshore deposits along the Cape south coast of South Africa.[1-7] These Pleistocene rocks have been the focus of a 350-km ground survey by a team led by C.W.H. between 2007 and 2019, from Arniston in the west to Robberg in the east. The resulting corpus of publications has dealt with mammal tracks and avian tracks. The only mention of reptile tracks was by Roberts et al.[1] who indicated that tracks of testudinae had been found in aeolianites east of Still Bay. However, no details were provided, and the site has not been re-located. Likewise, the fossil, archaeological and historical records have largely remained silent with regard to large terrestrial reptiles along this coastline, other than occasional reports of sea turtles, whereas there have been numerous reports on the mammalian fauna from archaeological sites, carnivore dens and scavenger dens.[8-10]

In 2018 our team identified palaeosurfaces containing tracks and trackways of large reptiles, and what we interpret as swim traces, in a coastal section of the Garden Route National Park (Figure 1). One of these surfaces also contained Middle Stone Age (MSA) lithics. The purpose of this article is to describe the tracks and track forms of these reptiles, consider plausible trackmakers, describe the lithics, and consider the palaeoenvironmental implications of these discoveries.

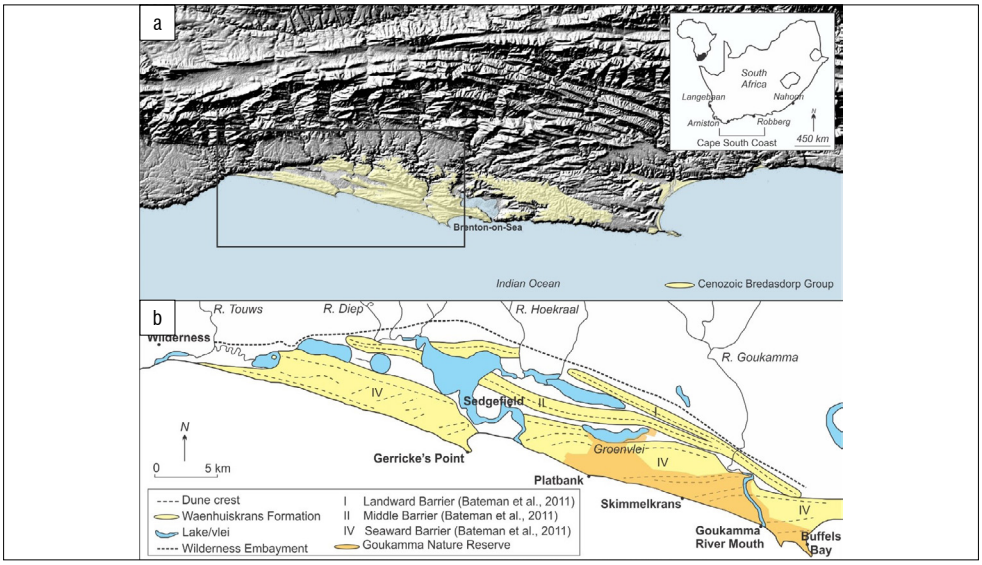


Figure 1: (a) Map of a portion of the Cape south coast showing sites mentioned in the text. (b) Enlarged area showing the Wilderness Embayment. The section of the Garden Route National Park containing reptile tracksites extends from Wilderness in the west to west of Platbank in the east.

Geological context

Pleistocene dunes and beaches occur as aeolianites and cemented foreshore deposits along the Cape south coast.[11-14] Aeolianite facies form the Waenhuiskrans Formation[15], and foreshore facies form part of the Klein Brak Formation, which Malan[16] interpreted as representing a succession of shallow marine, beach and estuarine or lagoonal deposits. These two formations form major elements of the Bredasdorp Group.[15,16] One of Malan's[16] stratotypes for the Klein Brak Formation is exposed at the Swartvlei estuary, situated east of the sites we describe.

The tracksites described lie within the Wilderness Embayment, the stratigraphic record of which has been described.[11,17,18] Bateman et al.[11] applied optically stimulated luminescence (OSL) dating to sites in this area: samples obtained from close to the tracksites we report on yielded a range of dates, the oldest being 148 ka ± 10 ka and the youngest being 92 ka ± 5 ka. Stratigraphy along this coastline consists mostly of composite Marine Isotope Stage (MIS) 6 and MIS 5 deposits, draped by a Holocene unconsolidated unit of dunes.[11] Around 91 ka, sea levels were as much as 45 m lower than at present, and the coastline in this area may have been as much as 60 km seaward of today's coast.[19] Around 126 ka, sea levels were 6.6–8 m higher than present levels on the Cape south coast.[20] Exposure of the continental shelf during sea-level lowstands created an environment characterised by lakes, lagoons, rivers and estuaries.[21]

Although palaeo-dunes generally dominate the Cape south coast compared to more rarely preserved beach deposits[12], Carr et al.[20] mapped out evidence of the MIS 5e highstand at Cape Agulhas, the Great Brak River mouth and Swartvlei Estuary. Further work was conducted at the Great Brak River mouth by Cawthra et al.[13], who described a regressive succession of shoreface, foreshore and aeolian deposits.

Methods

Once the tracks were identified in 2018, a search was conducted in the surrounding area for similar palaeosurfaces. Global Positioning Survey readings were recorded for all sites using a handheld device. Locality data were reposited with the African Centre for Coastal Palaeoscience and Garden Route National Park, to be made available to bona fide researchers upon request.

Measurements of tracks and traces included length, width, depth, pace length and stride length.[22] Results were recorded in centimetres. Expert trackers were invited to comment on tracks and potential trackmakers, through photographs and site visits. Tracings were made of selected tracks. Photographs were taken for photogrammetric analysis.[23] Three-dimensional 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 (v.2.10-beta).

As no in-situ tracksites were evident, in order to estimate the original angle of the track-bearing surfaces, we considered the orientation of rhizoliths in underlying layers, the position and size of displacement rims, and the degree of track asymmetry. While the sites could be visited during most tidal conditions, low tide proved optimal. Tracksites were examined at different times of day, to allow sunlight to optimally illuminate the surfaces. Standard field techniques were applied in understanding tracksite context. Dip and strike measurements were taken on primary beds, bed thicknesses were measured, and textures were described.

A site visit was undertaken to evaluate the stone artifacts. Preliminary observations and descriptions were made. As the artifacts were embedded in the palaeosurface, not all their aspects could be examined.

Results

None of the tracksites we describe were in situ; all occur on loose blocks and slabs of varying size. We were not able to conclusively identify the levels of origin of these rocks, but field correlation suggested that they may be located close to their original locations, near the foot of the coastal cliffs at or near the level of planar-bedded deposits that have the same primary sedimentary structures, approximately 10 m above the modern beach (Figure 2). As the cliffs were unstable and safe access was not feasible, this could not be confirmed.

In total, 14 track-bearing blocks and slabs were identified on the eastern and western margins of a headland. The westernmost and easternmost track-bearing surfaces were ~330 m apart. Some tracksites were on small slabs, containing just a single track. More information could be gleaned from larger surfaces, which typically contained multiple tracks.

A range of track sizes was apparent, with the largest tracks measuring ~22 cm in length and the smallest ~10 cm in length. All tracks were on natural mould surfaces, with the tracks forming impressions in the surface (concave epirelief). In some cases we concluded that the original surface had been close to horizontal, while in other cases it appeared to have been at an angle of up to 30°. Here we first present results from four larger surfaces, which we named Surfaces A, B, C and D.

Regardless of locality, all tracksites were preserved in comparable stratigraphy and lithology. Fine- to medium-grained, well-sorted sand that exhibited planar beds up to 20 cm in thickness, alternated with trough-cross bedding, dominated this geological unit. The vertical orientation of fallen blocks could be determined by the presence of the tracks. The surface stratum consisted of a thin (less than 2 cm) layer of silt and fine sand. This veneer of finer-grained material was apparently laterally

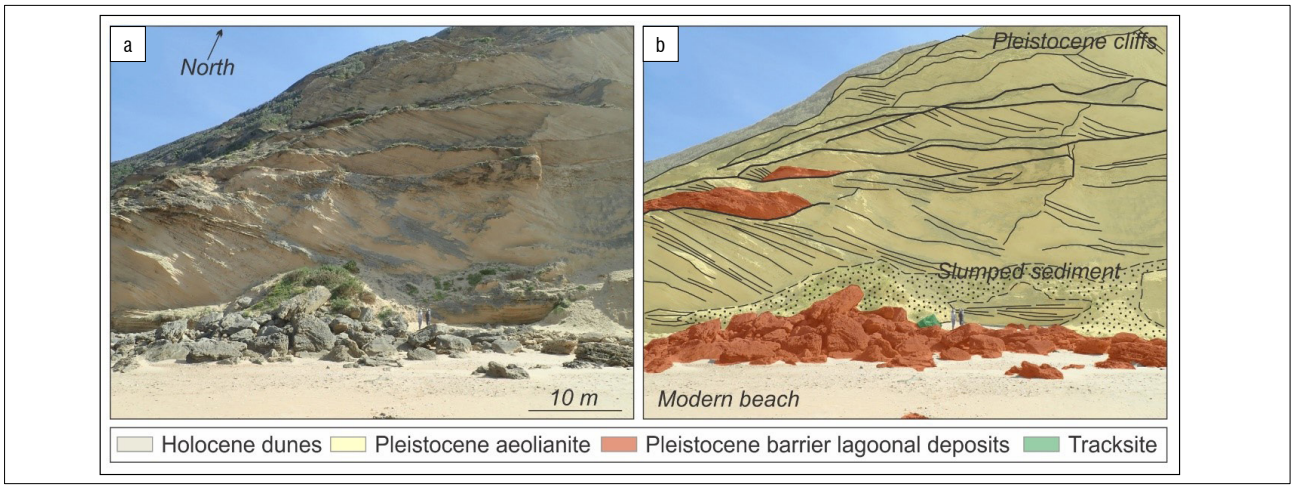


Figure 2: (a) Photograph of a trackway area, showing weathered aeolianite and fresh exposures where blocks have slumped onto the adjacent beach. (b) Marked and annotated section indicating a probable source area for the slumped blocks including the trackways. Aeolianite is illustrated in yellow, and planar, tabular beds interpreted to be remnants of a back-barrier lagoon are shown in orange. The tracksites are within the green area, beside two people who provide an indication of scale.

persistent as it was observed in association with tracks on both margins of the headland. The track-bearing layer inevitably occurred above and in contact with a heavily bioturbated layer up to 2 m in thickness. Rhizoliths occurred within these bioturbated sediments.

Surface A

The westernmost surface, which was inclined and east facing, measured 230 cm x 120 cm, on a loose rock 1.7 m in thickness (Figure 3a). Faint tracks could be discerned over much of the surface. Four large tracks were evident at the southern end (Figure 3b). Direction of motion was up the present slope of the surface. The two tracks on the left represented left-sided tracks, the two tracks on the right represented right-sided tracks, and the resulting trackway exhibited a wide straddle of 32 cm. The right upper (distal) track was orientated as much as 70° leftward compared to the other tracks. Substantial displacement rims were present posterolateral to the left proximal track and the right distal track.

The right proximal (lower) track was registered by the right forefoot. It was a partial track (being intersected by the bottom edge of the rock), and only showed the anterior ends of digital impressions.

The left proximal (lower) track was registered by the left hindfoot, and was characterised by a substantial displacement rim posterolaterally. Excluding the displacement rim but including the claw impressions, it was 22 cm long and 16 cm wide. Digit impressions were clearly evident; the longest was over 2.5 cm. Distance from the anterior edge of digits to the anterior end of the two claw impressions that were evident was 4 cm.

The left distal (upper) track was registered by the left forefoot, and was ~18–19 cm long and 16 cm wide. It exhibited at least five digit impressions. Digits II, III and IV were each ~7 cm long. Claw impressions were evident. The right distal (upper) track was made by the right hindfoot. It was less well preserved, but three deep digit depressions were evident, behind which were wide grooves, ending in a posterolateral displacement rim. Length could not be determined with precision; width was 16 cm.

The distance between hindfoot and forefoot (seen in the left tracks) was substantially less than the distance between forefoot and hindfoot (seen in the right tracks).

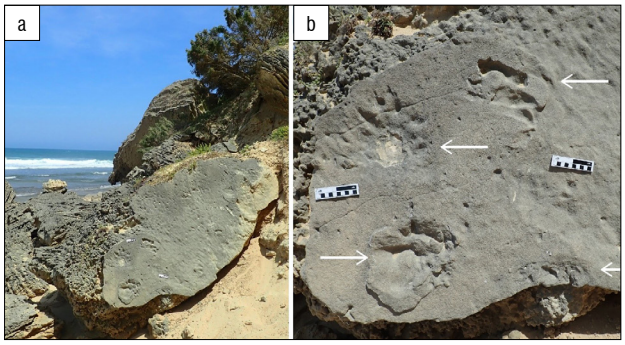


Figure 3: (a) Surface A. (b) Arrows indicate four large reptile tracks on Surface A; scale bars = 10 cm.

Surface B

Surface B lay 20 m east of Surface A, and was steeply inclined and west-facing. It measured 300 cm x 300 cm (Figure 4a). This surface presented a complex tracksite containing multiple reptile tracks and trackways, avian tracks, invertebrate burrow traces, and two stone artifacts.

The orientation of rhizoliths in the underlying bioturbated sediments suggested that the original surface had lain at an angle of ~30° (Figure 4b). This supposition was supported by the asymmetrical nature of tracks that crossed the surface, with upslope margins being steeper than downslope margins. Three trackways crossed the surface transversely, and isolated tracks were orientated in an upslope direction. Some of these formed composite tracks. A zone of parting lineation of flute casts covered the upper right portion of the surface. This veneer obscured some areas in which extensions of trackways might have been registered.

Despite the cornucopia of tracks, trackway interpretation was challenging: one of the trackways appeared partially covered by the above-mentioned sediment veneer, two trackways appeared to exhibit tracks heading in both directions (resulting in composite tracks), track size and pace lengths were not consistent, and some tracks were poorly defined. Nonetheless, some individual tracks displayed well-preserved features.

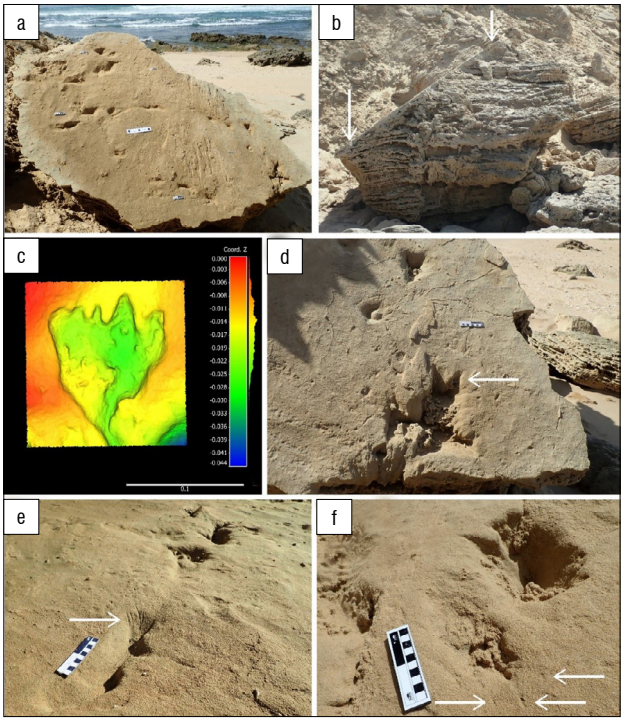


Figure 4: (a) Surface B; scale bars = 25 cm and 10 cm. (b) Arrows indicate Surface B in profile; bioturbated sediments and rhizoliths are evident. (c) Photogrammetry colour mesh of track on Surface B, showing parallel forward-pointing digits and possible erosional features following track registration. The three-dimensional model was generated with Agisoft MetaShape Professional (v. 1.0.4) using 27 images. Photos were taken an average 0.30 m from the surface. The reprojection error is 0.276 pix. Vertical and horizontal scales are in metres. (d) Arrows indicate deep digit impressions in large composite track on Surface B; scale bar = 10 cm. (e) Arrow indicates grooves in line with track in transverse trackway on Surface B; scale bar = 10 cm. (f) Arrows indicate claw marks ahead of small reptile track; scale bar = 10 cm.

One of the smaller depressions heading upslope was well preserved and showed what may have been four relatively parallel forward-pointing digits (Figure 4c). The posterior end tapered into a long, curved groove.

Deep digit impressions, similar to those seen on Surface A, and with a longer digit length (8 cm), were seen in one of the tracks heading upslope (Figure 4d). These were present at the distal end of a large, deep, composite impression. There were tantalising hints of parts of very large tracks at the bottom edge of the surface.

One of the tracks in the middle transverse trackway exhibited well-preserved narrow grooves that may represent claw scrape marks (Figure 4e). This trackway exhibited marked track asymmetry, with steep upslope margins and gentler downslope margins. The wide straddle typical of reptile tracks was not evident, and a reptile origin could therefore not be concluded.

In the upper transverse trackway, beside tracks heading in both directions, was a well-preserved smaller track, 17 cm long and 7 cm wide, with three small depressions (consistent with claw impressions) ahead of three of the digits (Figure 4f). The distance between the anterior ends of the digits and the anterior ends of the claw impressions ranged from 3 cm to 4 cm. In the upper right section, heading upslope, was a

small trackway containing three tracks, with four digits in an anisodactyl pattern consistent with an avian trackmaker, with tracks 3.0 cm long and 3.2 cm wide, and pace length of 13–14 cm.

The two stone artifacts (Figure 5) were embedded in the upper end of the surface, right of centre. They occurred beside the zone of parallel parting plane lineations.

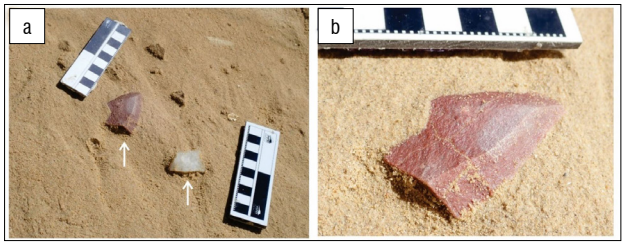


Figure 5: (a) Arrows indicate two stone artifacts on Surface B; scale bars = 10 cm. (b) Close-up of artifact on Surface B; scale bars in cm and mm.

The artifact (flake) to the left in Figure 5a (and Figure 5b), appeared to be composed of silcrete or quartzite; if it was the former, then its reddish hue suggested that it may have been heat treated. It measured 5.5 cm x 3.5 cm. The ventral surface and left edge (as viewed in Figure 5a) were embedded in the matrix. The right edge appeared to have some retouch. The butt of the striking platform, the proximal end on which the detachment blow or pressure had been applied, appeared to have broken off, possibly as potential preparation for hafting. The remaining medial edge of this flake was cracked; this cracking may have been due to internal flaws of the material or from heating of the silcrete. Negative scars were visible on the dorsal surface.

The artifact (flake) to the right in Figure 5a was of milky quartz material with a distal end embedded in the matrix. It appeared that the dorsal surface was exposed. It measured 5.0 cm x 3.0 cm.

The flakes were probably struck off prepared cores, although the milky quartz flake may be considered a small core. There were no chips or debitage on the surface to suggest that the flakes had been produced at the site.

Early in 2020, a substantial landslide buried Surface B. A nearby (non-reptile) tracksite was also buried in this event.

Surface C

The easternmost large surface, west-facing, measured 220 cm in height and 200 cm in maximum width. Rhizoliths were evident, most prominently in the bottom right portion. Much of the middle section exhibited scour marks and flute casts caused by water flow (from left to right) over a sand surface. At the top end were three deep, large impressions: a single upper impression, and two impressions in a lower row (Figure 6a).

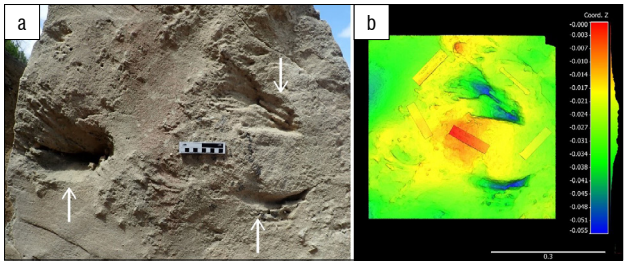


Figure 6: (a) Detail of a portion of Surface C, showing three deep depressions and curvilinear grooves, suggesting swim traces; scale bar = 10 cm. (b) Photogrammetry colour mesh: reptile swim traces on Surface C, using 62 images. Photos were taken an average 0.23 m from the surface. The reprojection error is 0.492 pix. Vertical and horizontal scales are in metres.

The two impressions in the lower row exhibited a characteristic, deep, curvilinear outline, 18 cm long, with convexity towards the bottom of the surface. To the left of each impression (as viewed in Figure 6a) were narrower curvilinear groove features, approximately 1 cm in diameter, ~2–3 cm apart, and 20–35 cm long (Figure 6b). These grooves extended through two of the larger depressions. If direction of movement was from right to left and the bottom row represented left tracks and the top impression was a right track, then these curvilinear grooves occurred anterior to the track impressions, and had a symmetrical appearance with outward convexity.

Surface D

Surface D lay 2 m to the west of Surface C. It was gently north-facing, measuring 215 cm x 60 cm (Figure 7a).

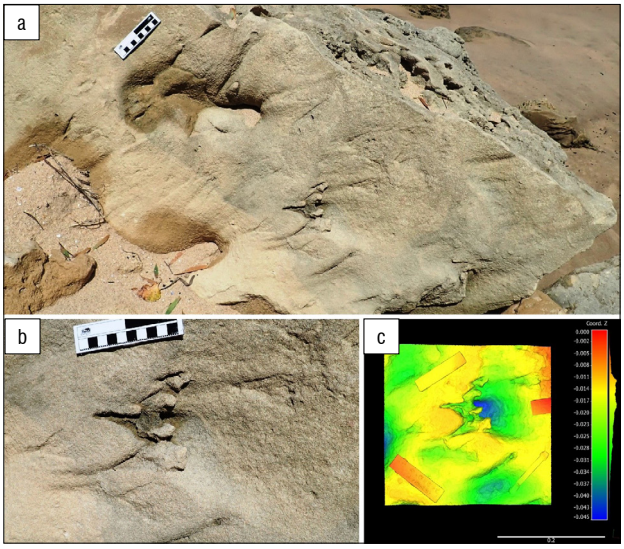


Figure 7: (a) Surface D; scale bar = 10 cm. (b) Detail on Surface D of reptile track and curvilinear grooves; scale bar = 10 cm. (c) Photogrammetry colour mesh: reptile tracks and swim traces on Surface D, using 42 images. Photos were taken an average 0.25 m from the surface. The reprojection error is 0.521 pix. Vertical and horizontal scales are in metres.

Surface D contained linear or curvilinear grooves, some of which were associated with digit impressions or larger, deeper impressions (Figure 7b, 7c).

The groove features were orientated in an approximately unimodal fashion, and were ~10 cm long. On a smaller surface beside Surface D were whitish cylindrical raised features (maximum dimensions 1.5 cm x 0.5 cm). Although rhizoliths and root casts occurred on nearby surfaces, and the shape of these features was not inconsistent with rhizoliths, macroscopically they appeared to contain osseous fragments. We therefore considered them more likely to be coprolites. They will be subjected to further analysis and will be reported on elsewhere.

Other sites

The remaining 10 sites confirmed that such tracks were a common local feature (Figure 8a–d); the majority of exposures of the distinctive track-bearing layer harboured reptile tracks. A small surface exposure situated between Surface A and Surface B contained a partial track with well-defined digit impressions (Figure 8a, 8b). We noted small reptile tracks (tortoise and perhaps terrapin) on two nearby surfaces with similar lithology. These will be described elsewhere.

Discussion

Track interpretation and trackmaker identification

The general crispness of the margins of the tracks is not consistent with a non-cohesive substrate such as soft, dry sand, and the depth of the tracks is not consistent with a very firm substrate.

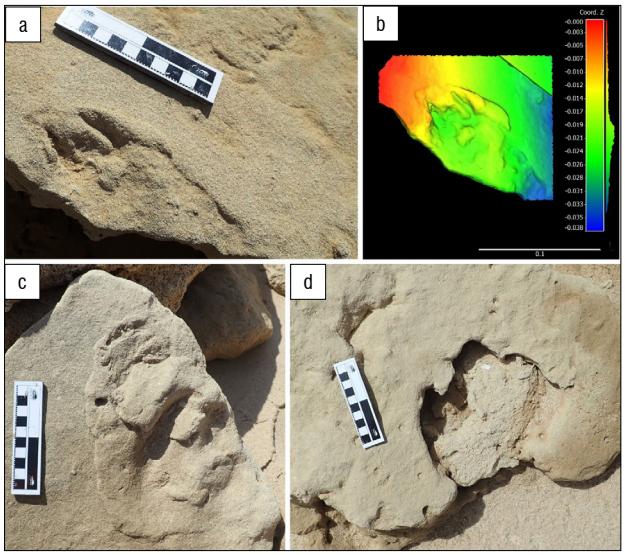


Figure 8: (a) Partial reptile track between Surface A and Surface B; scale bar = 10 cm. (b) Photogrammetry colour mesh: partial reptile track in (a), using 42 images. Photos were taken an average 0.33 m from the surface. The reprojection error is 0.398 pix. Vertical and horizontal scales are in metres. (c) Reptile track with large displacement rim; scale bar = 10 cm. (d) Probable large reptile track; scale bar = 10 cm.

The tracks appear to have been made on soft, wet substrates or underwater, in some cases on sloping surfaces, as suggested by the orientation of rhizoliths in underlying sediments. These factors may result in distortion of tracks, or sub-optimal preservation, compared with tracks made on firmer, drier surfaces. Suction effects may occur when a trackmaker's foot is removed from such a substrate, causing further distortion. This may explain the challenges we experienced in trackmaker identification despite the presence of many tracks. We interpreted the tracks of Surface A (Figure 3) as having been registered on a wet surface; the nature of the well-defined tracks, without evidence of swim traces, is not consistent with a deeper aqueous environment. This interpretation was buttressed by the presence of substantial displacement rims. The wide straddle, along with the distance between hindfoot and forefoot being less than that between forefoot and hindfoot, are consistent with the usual trackway pattern of a large reptile.[22,24,25] The more leftward orientation of the right upper (distal) track suggested a substantial swivel to the left. This interpretation was supported by the possible surfeit of digit impressions in the left upper (distal) track, which probably represented removal of the left forefoot from the surface, followed by its re-placement on the surface with a more leftward bearing. Furthermore, we interpreted the posterolateral displacement rims as being consistent with forward propulsion and a swivel to the left.

Surface B (Figure 4) represented an inclined surface; with evidence of having been wet or submerged when the tracks were registered. While we did not find conclusive evidence of tail drag impressions, we note that such impressions are less likely to be evident when tracks are registered on submerged surfaces. The zone of parting lineation of flute casts was consistent with water flow from the upper section towards the lower section. The sandstone veneer containing these flute casts, which obscured some portions of trackways, was thus younger than the track-bearing layer.

We interpreted the tracks on Surface C (Figure 6) as probably having been made on a submerged surface, based in part on the presence of the scour marks and flute casts that indicated the presence of water flow. The presence of possible swim traces supported this interpretation. We also interpreted the tracks on Surface D (Figure 7) as having been registered on a submerged surface; in this case the appearance of the curvilinear groove features, connecting with digit impressions or deeper impressions, suggested a relatively shallow aqueous environment.

We have considered extant southern African reptiles that arguably could have created such tracks and trackways. These include (1) the Nile crocodile (*Crocodylus niloticus*), which belongs to the family Crocodylidae, and is the only extant crocodylian (order Crocodylia) in southern Africa,[26] and (2) the two largest extant lizards in southern Africa, the water (Nile) monitor (*Varanus niloticus*) and rock monitor (*Varanus albigularis*), which belong to the order Squamata, family Varanidae.

During sea-level regressions in the terminal portion of the Middle Pleistocene and the Late Pleistocene, the vast Palaeo-Agulhas Plain was exposed along the Cape south coast.[27] It was characterised by extensive floodplains and wetlands,[19,28] and a distinctive megafauna of which a number of species are extinct. These wetlands may have provided habitat for large reptiles, although the warm conditions typically associated with large reptiles[26] may not have been present. In assessing the track evidence, we also need to consider the possibility, albeit remote, that tracks may have been registered by an extinct, hitherto undescribed reptile.

The height of the possible source horizon is above present sea level; based on this height, as well as the chronology of the coastal cliff sequences,[11] we posit an age of MIS 5e. Wetland environments, lagoons and back-barrier systems are well preserved in the highstand record as well as on the continental shelf, with their facies commonly identified in Pleistocene coastal cliffs.[1,11,13,20] Our field evidence suggests that these trackways were mostly made in a sub-aqueous environment – interpreted to have been a back-barrier lagoon or interdune lake. The silt-rich stratum in which the tracks are preserved suggests depositional quiescence which, in coastal environments, is representative of a low-energy system such as a lagoon or interdune lake.[29]

The track and trackway morphology (digit shape, digit size, claw impressions, number of digits, wide straddle) provide evidence of large reptile trackmakers.[22,24,25] The challenge is to determine whether these trackmakers can be identified to family or genus level, and whether or not more than one trackmaker species may have created the tracks. We do not believe that differences in track morphology between the two *Varanus* species are substantial enough to allow us to distinguish between them if they occur on these palaeosurfaces.[24,25] However, *V. albigularis* would be a less likely trackmaker in an aquatic environment. The differences between the tracks of *C. niloticus* and *V. niloticus* are more substantial, including shape and orientation of the digits and distance between the anterior end of the toe impressions and the claw impressions (typically greater in *Varanus*).[24,25]

Using a formula devised by Thomson[30] for estimating the length of *C. niloticus* based on the size of the hindfoot, and applying this to Surface A, a value of 257 cm was obtained. We are not aware of a similar formula for estimating length of *Varanus* species. We do not use track size as a sole criterion, because post-depositional factors could conceivably affect track size, because reptile growth is incremental over an animal's life span, because *V. niloticus* can reach large sizes, and because the reported size of extant reptiles may not be the same as the size attained in the Pleistocene: Klein[31] reported that mammalian carnivore size varied significantly during the Pleistocene, being larger during glacial phases, and Tyrberg[32] reported that at a global level Late Pleistocene bird species were often larger than their extant descendants. The same phenomenon might apply to reptiles. Nonetheless, 257 cm is greater than the 220 cm upper limit of the *V. niloticus* size range reported by Keates[33]. The digit impressions of the large partial track on Surface B are longer than those measured at Surface A. While this may seem to suggest an even larger trackmaker, differences in substrate can also account for such variations. *V. albigularis* is smaller than *V. niloticus*, and is even less likely to have attained such a large size. The expert trackers whom we consulted concluded that the large tracks evident on Surface A were consistent with those of *C. niloticus*. Based on track dimensions, digit shapes, and claw impressions, we conclude that the large tracks on Surface A are consistent with those of the extant *C. niloticus* or a similar large reptile.

We are not aware of previous reports of fossil reptile swim traces from Africa, although hippopotamus swim traces have been reported.[34] We are

not aware of reports of fossil non-crocodylian reptile swim traces. However, crocodylian swim traces feature in the global ichnological record.[35,36] There is a range of swim trace morphologies, depending on depth of water (and the size and leg length of the swimmer), with tracks made in shallow water at one end of this spectrum (virtually indistinguishable from dry-land tracks), and faint, slightly curvilinear impressions made in deep water at the other end. Traces made in relatively deep water are described as 'raking traces', forming parallel or sub-parallel scratch marks.[37] Clearly a reptile that is exclusively swimming, and not touching the bottom, leaves no traces whatsoever. Examination of swim traces of extant C. *niloticus* and *V. niloticus* would be useful to further document the nature of this spectrum. The flippers of large adult turtles tend to strike the substrate more laterally, and would therefore leave impressions unlike those seen in Figure 6; hence such species can probably be excluded. We interpret the features at Surface C and Surface D as including probable reptile swim traces, while acknowledging that some of the features on Surface C could conceivably be tail drag impressions. Some of the features on Surface C appear consistent with 'raking traces', and therefore appear to have been made in a deeper aqueous environment than those on Surface D, where shorter groove features are contiguous with well-defined digit impressions.

Despite the many smaller tracks and probable swim traces we observed, determining trackmaker identity with confidence proved challenging. Two exceptions were apparent. The partial track shown in Figure 8a and 8b resembles that of a juvenile *C. niloticus*. The track shown in Figure 4f exhibits a substantial, consistent distance between the anterior ends of the digits and the claw impressions, which led us to conclude that the trackmaker may have been a lizard species similar to the two extant *Varanus s*pecies.

The fossil and archaeological record

Pickford[38] described crocodile remains from Early and Middle Miocene deposits in Namibia. Hendey[39] described what is possibly the earliest known occurrence of *C. niloticus* from Middle Miocene deposits at Arrisdrift, on the Namibian side of the Orange River. Feely[40] provided a review of known southern occurrences of *C. niloticus*, and its current, historical and prehistorical southern range limits. He noted a skull specimen of an unknown saurian found by A.G. Bain in the Eastern Cape at a place Bain called 'Crocodile Hill'. However, the site and geological context are unknown. A large, partially fossilised skull of *C. niloticus*, of uncertain age, was recovered near the Swartkops estuary east of Port Elizabeth.[40]

Closer to our study area are two further possible records.[40] One is a report of a crocodile tooth found in a shelter at Lottering River (just within the Eastern Cape Province), presumed to have been transported there by a Later Stone Age inhabitant. The second is a report of a crocodile skull from Matjes River, east of Plettenberg Bay, also presumed to have been transported by a Later Stone Age inhabitant. The archaeological context is unknown in both cases.

There is an absence of reports of large reptiles from archaeological sites or contemporary faunal accumulations on the Cape south coast; Richard Klein (2019, written communication, February 15) indicates that in his extensive inventory of the region there are no coastal records of crocodiles or monitor lizards. Literature on Pleistocene reptiles from the Cape south coast from archaeological sites relates to small reptiles (tortoises).[41,42]

Historical records, folklore and current range limits

West of the tracksites we describe, the Kaaimans River flows into the Indian Ocean. Leggatt[43] has researched the history of this name, which could signify 'cayman'. He notes that explorers' accounts from the 18th century use the name with varying spelling, that Le Vaillant in 1782 associated it with crocodiles, and that Lichtenstein in 1803 thought it referred to the water monitor.[43] However, these associations were presented without evidence. Leggatt[43] provides an alternative derivation, from a Khoisan belief in a giant watersnake, 'Keiman', that inhabited river pools. Feely[40] reaches a similar conclusion, while also noting that the name might derive from a smaller lizard.

There are no known historical records of large reptiles from the Cape south coast, nor from adjoining areas. This absence of evidence includes accounts of the coastal regions of the Western Cape and Eastern Cape by survivors of Portuguese shipwrecks between Plettenberg Bay and Pondoland in the 16th and 17th centuries.[40]

Due to its ectothermic biology, *C. niloticus*, like most other crocodylians, is limited to tropical and subtropical areas with daily maximum ambient temperatures of 25–35 °C.[44] The southward cooling of habitats forms the most significant determinant of its range limits. All recorded localities for this species towards the southern limits of its range are at or near coastal estuaries or wetlands. The closest *C. niloticus* population to the tracksites described here is at the Dwesa-Cwebe Nature Reserve in the Eastern Cape Province.[40,45] Six juvenile *C. niloticus* were introduced from Zululand to this area in 1977,[45] and two successful breeding events have been documented (Venter J 2014, written communication).

Less information is available on the historical or current distribution of *V. niloticus*, although similar considerations can be assumed regarding the southward cooling of temperature. Lubke and De Moor[46] report that it extends southwest as far as Kouga River, Eastern Cape. Keates[33] reports the southern range limit at Seekoei River, Eastern Cape. Both localities are more than 200 km east of the tracksites we describe.

Lithics associated with fossil tracks

Although observations of the artifacts on Surface B are necessarily preliminary (Figure 5), their characteristics suggest that they date from the MSA.[47-50] We are not aware of prior reports of lithics embedded in Pleistocene palaeosurfaces from the Cape south coast, and certainly none that occur in association with fossil tracks. Roberts et al.[1] reported 'six non-diagnostic stone artefacts' from a palaeosol between aeolianite layers east of Still Bay. Graham Avery (2019, written communication, February 10) reports occurrences of artifacts in association with aeolianite outcrops from False Bay (at the Holocene / Late Pleistocene erosional surface in the Olympic Sand Mine). From the West Coast, Cruz-Uribe et al.[51] reported Late Acheulean artifacts in palaeosols between calcareous deposits dating to as much as 160 ka at Duinefontain, and Avery et al.[52] reported artifacts and shell middens in an eroded overhang dated to MIS 5 at Ysterfontein. To the east, Kevin Cole (2019, written communication, February 5) reports blades and scrapers embedded in an aeolianite surface, near the Nahoon Point hominin tracks, and Graham Avery (2019, written communication, February 10) reports having found a MSA convergent flake in this same area. Pargeter et al.[53] described lithics in Middle Pleistocene dune surfaces from the Pondoland coast. The significance of such occurrences may be greater if they are found on the same surfaces as fossil tracks.

We found no evidence to suggest that any tracks on Surface B were of hominin origin, and the lithics were embedded in the layer of sediment that appeared to have washed over the track-bearing surface. Nonetheless, a spatial and temporal association in this environment between humans and large reptiles can be inferred, or at least mutual use of habitat, something which has not previously been documented. The MSA date for the lithics is consistent with the range of OSL dates obtained from nearby deposits.

The loss of this site to further study as a result of a landslide is regrettable. However, this is consistent with the ephemeral nature of many tracksites on the Cape south coast. It underlines the importance of early detection and documentation of newly exposed tracksites.

Palaeoenvironmental considerations

For *C. niloticus* embryos to hatch successfully, incubation (i.e. sand) temperatures must be sufficiently warm (28–35 °C) for egg incubation.[54] Measurement of sand temperatures close to the sites we describe would establish whether or not egg incubation is currently theoretically possible on the Cape south coast. We suspect that it is not, and that the reptile tracks were registered during a warmer period during the Pleistocene during or close to the time of a sea highstand that allowed reptile species to occupy an extended range. Chronological context[11] suggests an age of MIS 5e (the Last Interglacial). As sea levels during MIS 5e in this area were up to 6–8 m higher than at present, a warmer climate capable of supporting large reptiles on the Cape south coast can be inferred.[20]

Conclusions

The sites we describe represent the first evidence of large reptiles (such as *C. niloticus* and *V. niloticus*) on the Cape south coast of South Africa, and probably the first recorded reptile swim traces in Africa. They join the substantial record of avian and mammal Pleistocene tracksites along this coastline, and complement the extensive record of body fossils. Their importance includes palaeoenvironmental implications, as they were probably registered during a sea highstand characterised by warm temperatures – consistent with the ectothermic biological requirements of large reptiles such as *C. niloticus* and probably *V. niloticus*. The finding of two MSA artifacts on one of the track-bearing surfaces represents a previously undocumented association of Pleistocene hominin populations with large reptiles. Further studies could include recovery and more detailed analysis of the artifacts, measurement of sand temperatures close to the tracksites, and documentation of the swim traces of extant large African reptiles.

Acknowledgements

We acknowledge the support of Graham Avery, Andre Brink, Emily Brink, Kevin Cole, Jessica Hayes and staff of Garden Route National Park, Linda Helm, Richard Klein, Martin Lockley, Liezl Van Pletzen-Vos and Jan Venter. We thank expert trackers Ian Thomas and Alan Yeowart for their valuable opinions.

Authors' contributions

C.W.H.: Lead author, corresponding author, conceptualisation, data collection, data analysis, writing, project leadership, track analysis. H.C.C.: Conceptualisation, sample analysis, site visit, data analysis, contribution on geological context, field stratigraphy, review of drafts and revisions. X.C.: Conceptualisation, data analysis, site visit, contributions on crocodile records and reptile biology, review of drafts and revisions. C.J.Z.H.: Conceptualisation, data analysis, photogrammetry, review of drafts and revisions. R.R.: Conceptualisation, data analysis, site visit, contributions on MSA lithics, review of drafts and revisions. W.S.: Conceptualisation, data analysis, site visits, contributions on sedimentology, review of drafts and revisions. A.v.d.H.: Conceptualisation, data analysis, site visit, contributions on ichnology and track analysis, review of drafts and revisions.

References

1. Roberts DL, Bateman MD, Murray-Wallace V, 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
2. Helm CW, Anderson RJ, Buckley LG, Cawthra HC, De Vynck JC. Biofilm assists recognition of avian trackways in Late Pleistocene coastal aeolianites, South Africa. Palaeont Afr. 2017;52:78–84.
3. 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
4. Helm CW, McCrea RT, Cawthra HC, Thesen GHH, Mwankunda JM. Late Pleistocene trace fossils in the Goukamma Nature Reserve, Cape south coast, South Africa. Palaeont Afr. 2018;52:89–101.
5. 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
6. 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.
7. Helm CW, Cawthra HC, De Vynck JC, Lockley MG, McCrea RT, Venter J. The Pleistocene fauna of the Cape south coast revealed through ichnology at two localities. S Afr J Sci. 2019;115(1/2), Art. #5135, 9 pages. https://doi.org/10.17159/sajs.2019/5135
8. Klein RG. The late quaternary mammalian fauna of Nelson Bay Cave (Cape Province, South Africa): Its implications for megafaunal extinctions and environmental and cultural change. Quat Res. 1972;2:135–142. https://doi.org/10.1016/0033-5894(72)90034-8
9. Klein RG. The mammalian fauna of the Klasies River Mouth sites, southern Cape Province, South Africa. S Afr Archaeol Bull. 1976;31:75–98. https://doi.org/10.2307/3887730
10. Klein RG. Palaeoenvironmental implications of Quaternary large mammals in the Fynbos Region. In: Deacon HJ, Hendey QB, Lambrechts JJN, editors. Fynbos palaeoecology: A preliminary synthesis. Pretoria: Cooperative Scientific Programmes; 1983. p. 116–138.
11. 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
12. 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
13. 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
14. 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
15. 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.
16. 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.
17. 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
18. 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
19. 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
20. 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
21. Cawthra HC, Cowling RM, Andò S, Marean CW. Geological and soil maps of the Palaeo-Agulhas Plain for the Last Glacial Maximum. Quat Sci Res. In press 2019. https://doi.org/10.1016/j.quascirev.2019.07.040
22. Stuart C, Stuart T. A field guide to the tracks and signs of southern and East African wildlife. Cape Town: Struik Nature; 2000.
23. Matthews NA, Noble TA, Breithaupt BH. Close-range photogrammetry for 3-D ichnology: The basics of photogrammetric ichnology. In: Falkingham PL, Marty D, Richter A, editors. Dinosaur tracks: The next steps. Bloomington, IN: Indiana University Press; 2016. p. 28–55.
24. Liebenberg L. A photographic guide to tracks and tracking in southern Africa. Cape Town: Struik Publishers; 2000.
25. Van den Heever A, Mhlongo R, Benadie K. Tracker manual – A practical guide to animal tracking in southern Africa. Cape Town: Struik Nature; 2017.
26. Branch B. Field guide to the snakes and other reptiles of southern Africa. Cape Town: Struik; 1994.

27. 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

28. Cawthra HC. The marine geology of Mossel Bay [PhD thesis]. Cape Town: University of Cape Town; 2014.

29. Nichols G. Sedimentology and stratigraphy. Chichester; John Wiley and Sons Ltd.; 2009.

30. Thomson WR. A field guide for estimating the overall length of the crocodile. S Afr J Wildl Res. 1972;2(1):27.

31. Klein RG. Carnivore size and Quaternary climatic change in southern Africa. Quat Res. 1986;26:153–170. https://doi.org/10.1016/0033-5894(86)90089-X

32. Tyrberg T. The Late Pleistocene continental avian extinction – An evaluation of the fossil evidence. Oryctos. 2007;7:249–269.

33. Keates C. The monitor lizards of southern Africa [webpage on the Internet]. c2016 [cited 2019 Nov 08]. Available from: https://nextgenherpetologist.co.za/2016/12/23/the-monitor-lizards-of-southern-africa/

34. Bennett MR, Morse SA, Falkingham PL. Tracks made by swimming hippopotami: An example from Koobi Fora (Turkana Basin, Kenya). Palaeogeogr Palaeoclimatol Palaeoecol. 2014;409:9–23. http://dx.doi.org/10.1016/j.palaeo.2014.04.021

35. Lockley MG, Lucas SG, Milàn J, Harris JD, Avanzini M, Foster JR, et al. The fossil record of crocodylian tracks and traces: An overview. In: Milàn J, Lucas SG, Lockley MG, Spielmann JA, editors. Crocodyle tracks and traces. New Mexico Museum of Natural History and Science Bulletin 51. Albequerque, NM: New Mexico Museum of Natural History and Science; 2010. p. 1–15.

36. Lockley MG, Lucas SG. Fossil footprints of western North America. New Mexico Museum of Natural History and Science Bulletin 62. Albequerque, NM: New Mexico Museum of Natural History and Science; 2014.

37. Romano M, Whyte MA. Crocodilian and other non-dinosaurian tracks and trackways from the Ravenscar Group (Middle Jurassic) of the Cleveland Basin, Yorkshire, UK. In: Milàn J, Lucas SG, Lockley MG, Spielmann JA, editors. Crocodyle tracks and traces. New Mexico Museum of Natural History and Science Bulletin 51. Albequerque, NM: New Mexico Museum of Natural History and Science; 2010. p. 69–81.

38. Pickford M. A new species of crocodile from Early and Middle Miocene deposits in the lower Orange River Valley, Namibia and the origins of the Nile crocodile (*Crocodylus niloticus*). In: Pickford M, Senut B, editors. Memoir of the Geological Survey of Namibia: Geology and Palaeobiology of the Central and Southern Namib: Palaeontology of the Orange River Valley, Namibia. Volume 19. Windhoek: Geological Survey of Namibia; 2003. p. 51–65.

39. Hendey QB. Preliminary report on the Miocene vertebrates from Arrisdrift, South West Africa. Ann S Afr Mus. 1978;76:1–41.

40. Feely JM. On the southeastern range limits of the Nile crocodile: A review of its past and present occurrences in the Eastern Cape and Western Cape, South Africa. S Afr J Wildl Res. 2010;40(2):169–175. https://doi.org/10.3957/056.040.0209

41. Reynard JP, Henshilwood CS. Subsistence strategies during the Late Pleistocene in the southern Cape of South Africa: Comparing the Still Bay of Blombos Cave with the Howiesons Poort of Klipdrift Shelter. J Hum Evol. 2017;108:110–130. https://doi.org/10.1016/j.jhevol.2017.04.003

42. Thompson JC, Henshilwood CS. Tortoise taphonomy and tortoise butchery patterns at Blombos Cave, South Africa. J Archaeol Sci. 2014;41:214–229. https://doi.org/10.1016/j.jas.2013.08.017

43. Leggatt H. Wilderness names [document on the Internet]. c2016 [cited 2019 Nov 08]. Available from: http://wrra.co.za/images/stories/PDF/HugoLeggatt/Wilderness%20Names.pdf

44. Grigg G, Kirshner D. Biology and evolution of crocodylians. New York: Cornell University Press; 2015.

45. Pooley AC. Crocodile research in Maputaland. In: Bruton MN, Cooper KH, editors. The ecology of Maputaland. Cape Town: Cape & Transvaal Printers (Pty) Ltd; 1980. p. 293–299.

46. Lubke R, De Moor I. Guide to the eastern and southern Cape coasts. Cape Town: UCT Press; 1998.

47. Wurz S. Middle Stone Age tools from Klasies River main site, conventions and symbolic cognition. In: Nowell A, Davidson I, editors. The cutting edge: Stone tools and the evolution of cognition. Boulder, CO: Colorado University Press; 2010. p. 135–158.

48. Wurz S. New investigations at Klasies River Main site. The Digging Stick. 2016;33(3):7–10.

49. 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

50. Lombard M. Evidence for hunting and hafting at Sibudu Cave during the Middle Stone Age. The Digging Stick. 2002;19(3):1–3.

51. Cruz-Uribe K, Klein RG, Avery G, Avery DM, Halkett D, Hart T, et al. Excavation of buried Late Acheulean (Mid-Quaternary) land surfaces at Duinefontein 2, Western Cape Province, South Africa. J Archaeol Sci. 2003;30(5):559–575. https://doi.org/10.1016/S0305-4403(02)00202-9

52. Avery G, Halkett D, Orton JDJ, Steele TE, Tusenius M, Klein RG. The Ysterfontein 1 Middle Stone Age rock shelter and the evolution of coastal foraging. Goodwin Ser. 2007;10:66–89. https://www.jstor.org/stable/40650020

53. Pargeter J, Saktura R, Blackwood A, Cawthra H, Esteban I, Fisher E, et al. The Msikaba Red Sand Dunes: Middle Pleistocene lithic technological variability in Pondoland, South Africa. Paper presented at: The 84th Annual Meeting of the Society for American Archaeology; 2019 April 10–14; Albequerque, New Mexico, USA.

54. Hutton JM. Incubation temperatures, sex ratios and sex determination in a population of Nile crocodiles (*Crocodylus niloticus*). J Zool. 1987;211:143–155. https://doi.org/10.1111/j.1469-7998.1987.tb07458.x

AUTHORS:
Dominic Stratford[1]
Laurent Bruxelles[1,2,3]
J. Francis Thackeray[4]
Travis R. Pickering[5]
Sophie Verheyden[6]

AFFILIATIONS:
[1]Department of Archaeology, School of Geography, Archaeology and Environmental Studies, University of the Witwatersrand, Johannesburg, South Africa
[2]French National Institute for Preventive Archaeological Research, French Institute of South Africa (IFAS), Johannesburg, South Africa
[3]Sub-Saharan Africa (USR 3336), French National Center for Scientific Research (CNRS), Paris, France
[4]Evolutionary Studies Institute, School of Geoscience, University of the Witwatersrand, Johannesburg, South Africa
[5]Department of Anthropology, University of Wisconsin-Madison, Madison, Wisconsin, USA
[6]Royal Belgian Institute of Natural Sciences, Brussels, Belgium

CORRESPONDENCE TO:
Dominic Stratford

EMAIL:
Dominic.Stratford@wits.ac.za

DATES:
Received: 09 Sep. 2019
Revised: 06 Jan. 2020
Accepted: 07 Jan. 2020
Published: 26 Mar. 2020

HOW TO CITE:
Stratford D, Bruxelles L, Thackeray JF, Pickering TR, Verheyden S. Comments on 'U-Pb dated flowstones restrict South African early hominin record to dry climate phases' (Pickering et al. Nature 2018;565:226–229). S Afr J Sci. 2020;116(3/4), Art. #7094, 2 pages. https://doi.org/10.17159/sajs.2020/7094

ARTICLE INCLUDES:
☒ Peer review
☐ Supplementary material

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

EDITOR:
Maryna Steyn

KEYWORDS:
palaeoanthropology, archaeology, palaeoenvironment, karst science, geomorphology

FUNDING:
None

Comments on 'U-Pb dated flowstones restrict South African early hominin record to dry climate phases' (Pickering et al. Nature 2018;565:226–229)

Pickering et al. (Nature 2018;565:226–229) utilised calcium carbonate flowstone deposits (i.e. speleothems) from eight Pliocene and Pleistocene South African Cradle of Humankind cave sites to propose that biases were created within the fossil record due to absent clastic sedimentation phases during wet periods, when caves were closed and only speleothems accumulated. Such a scenario has significant implications for our understanding of variability in hominin mobility, resource exploitation, functional repertoires and interactions with competitors in changing environmental and ecological contexts. We find considerable issues with the article. First, Pickering et al.'s contribution omits crucial fossil evidence from various stratigraphic units of the Sterkfontein Caves that indicates conditions were not always arid when the caves were open and sediments were deposited. Second, Pickering et al.'s proposal that clastic and speleothemic deposits (including faunal and floral material) form mutually exclusively is an overly simplified, binary depositional (and in this case environmental) framework that demonstrates an inherent bias in the sampling of cave deposits for dating. This creates the impression that either speleothems or clastic sediments are deposited and does not take into account the full spectrum of sedimentary complexity in karst caves. Third, closure of the caves across the Cradle of Humankind landscape during wet periods is not substantiated geomorphologically or speleologically; identification of the responsible process is critical to the proposed infilling scenario.

Significance:

- We propose that Pickering et al.'s interpretation of the environmental context of the South African early hominin record is problematic in that it omits crucial faunal and floral fossil evidence associating hominins with non-arid climates, is geomorphologically unsupported, and perpetuates biases against temporally and climatically representative clastic sediments due to challenges related to their dating.

Pickering et al. (2018)[1] utilise calcium carbonate flowstone deposits from eight Pliocene and Pleistocene South African Cradle of Humankind (also referred to as 'the Cradle') cave sites to propose that biases are created within the fossil record due to absent clastic sedimentation phases during wet periods. They propose a close correlation between cave closure, flowstone growth and 'phases of increased effective precipitation', suggesting that clastic sedimentation and calcium carbonate precipitation are mutually exclusive processes, with clastic deposits (including faunal and floral material) forming only during arid periods. This scenario has significant implications for our understanding of variability in hominin mobility, resource exploitation, functional repertoires and interactions with predators in changing environmental and ecological contexts. We find this proposition to be problematic as a result of oversimplification. Reducing environmental conditions to a 'binary' framework of such extreme end-member states of wet or dry does not realistically reflect the majority of time represented by intermediate, moderately variable climatic conditions during which deposits also accumulate. Reducing sediment accumulation to one end of the climatic spectrum limits the validity of nuanced interpretations of hominin ecological relationships represented by the diverse fossil assemblages interred in the extensive clastic deposits. Specifically, we find the following issues with the article:

1. Published evidence clearly associating hominins in part with wet ecological conditions, was not considered seriously enough in the article. For example, the authors omit crucial fossil evidence from the Sterkfontein Caves of liana vines that indicate the presence of ancient dense woodland or forest at ~2.5 million years ago.[2] The authors also omit important faunal data, notably the presence of fossil colobine monkeys, Alcelaphini bovids such as *Damaliscus* sp. and *Megalotragus* sp. (which are dependent on savanna grassland), and other broken-open country antelopes like *Tragelaphus* sp. and *Antidorcas* from various Sterkfontein stratigraphic levels.[3-6] These reflect at least some degree of savanna woodland, which, in turn, contradicts inferences of dry conditions. Similarly, micromammal samples from Sterkfontein[7] demonstrate a mosaic environment (that importantly include moist, woodland settings as suggested by the presence of *Elephantulus fuscus* fossils), and represent interglacial (comparatively warm and wet) conditions which prevailed during sedimentation and fossil accumulation. Further, stable carbon isotope data derived from tooth enamel of a range of taxa, including hominins[8-10], relate to diets that reflect consumption of a significant amount of C3 vegetation, indicating, in turn, that conditions were certainly not always arid at times when the caves were open and sediments were being deposited.
2. Generally, dates for clastic sediment deposits are ascertained through the application of U-Pb or palaeomagnetic dating of interstratifying (or assumed to be interstratifying) flowstones.[11] The clastic sediments (collectively known as breccias) themselves are not often suitable for dating using the comparatively broadly applied palaeomagnetic seriation and U-Pb methods.[12-14] Consequently, often only speleothems are sampled, creating the impression that speleothem or clastic sediments are mutually exclusively deposited. Clastic sediments deposited nearby at the same time will largely remain undated and therefore overlooked.
3. Pickering et al.'s proposal of Cradle-wide cave closure during wet environmental conditions is problematic from both karst geological and geomorphological perspectives. In many contemporary situations around

the world, flowstones form in open caves that are also actively accumulating sediments, for example: Aven d'Orgnac, France; Dadong Cave, southern China; Xkeken Cenote, Mexico; Kotilola Cave, Papua New Guinea; Cave of the Owls, Peru; Son Doong Cave, Vietnam; and Mbannza-Ngungu Caves, Democratic Republic of the Congo. Moreover, there is convincing evidence that coeval or penecontemporaneous clastic and speleothem formation occurred in ancient[12,15,16] and archaeological[17,18] deposits. Pickering et al. do not explain the geomorphological mechanisms that bind cave closing processes to wet periods. However, this physical process is important to identify and apply across the whole Cradle of Humankind karstic landscape, especially given the identified implications for landscape-wide cave closure during specific environmental conditions. The diversity of geological and geomorphological contexts across the Cradle of Humankind landscape[19] makes application of a uniform Cradle-scale cave closure process challenging and in this case unsupported. While speleothem deposition seems to have a strong relationship with climate, more specifically with water and CO_2 availability[20], the opening and closing of a cave system seems to be primarily determined by erosional cycles (and tectonic changes), and is therefore potentially strongly influenced at the local scale and only indirectly influenced by climate. While little literature exists on the mechanisms of cave entrance opening and closure in palaeokarst, authors describe vadose zone openings as enlarging during wetter periods due to increased recharge flow rates.[21,22] These processes result in a complex succession of sedimentation processes[23] and sedimentary deposition and speleothem deposition should not be considered mutually exclusive phenomena.

In summation, faunal and floral evidence clearly indicate that hominin fossils are not necessarily always associated with arid climates and that fauna and flora accumulate in open caves through a broad range of environmental conditions. Pickering et al.'s methodology for sampling cave deposits for dating, which focuses on the application of U-Pb or palaeomagnetic techniques to interstratifying (or assumed to be interstratifying) flowstones, produces ages (and by proxy cave sedimentation conditions) only for speleothems. The proposed correlation of speleothem formation with cave closure creates a convenient but largely unsupported deterministic interpretation of palaeoenvironmental conditions that necessarily excludes the significant clastic deposit record. Their resulting interpretation of the environmental context of the South African hominin record consequently perpetuates biases against temporally and climatically representative clastic sediments. Ideally, palaeoenvironmental reconstructions should be developed that integrate clastic and speleothemic evidence in coherent sequences prior to isolating and excluding evidence.

Acknowledgements

We thank the anonymous reviewers for their valuable input which helped us to improve this short contribution. D.S. thanks Aurore Val and Jennifer Fitchett for their comments and the editors of SAJS.

Authors' contributions

All authors contributed equally to the conceptualisation, data presentation and editing of the text.

References

1. Pickering R, Herries AIR, Woodhead JD, Hellstroom JC, Green HE, Paul B, et al. U-Pb dated flowstones restrict South African early hominin record to dry climate phases. Nature. 2018;565:226–229. https://doi.org/10.1038/s41586-018-0711-0
2. Bamford M. Pliocene fossil woods from an early hominid cave deposit, Sterkfontein, South Africa. S Afr J Sci. 1999;95:231–237.
3. Vrba ES. Some evidence of chronology and palaeoecology of Sterkfontein, Swartkrans and Kromdraai from the fossil Bovidae. Nature. 1975;254:301–304. https://doi.org/10.1038/254301a0
4. Vrba ES. The fossil Bovidae of Sterkfontein, Swartkrans and Kromdraai. Transvaal Museum Memoir 21. Pretoria: Transvaal Museum; 1976.
5. Brain CK. The hunters or the hunted: An introduction to African cave taphonomy. Chicago, IL: University of Chicago Press; 1981.
6. Luyt CJ. Revisiting palaeoenvironments from the hominid-bearing Plio-Pleistocene sites: New isotopic evidence from Sterkfontein [MSc thesis]. Cape Town: University of Cape Town; 2001.
7. Avery DM. The Plio-Pleistocene vegetation and climate of Sterkfontein and Swartkrans, South Africa, based on micromammals. J Hum Evol. 2001;41:113–132. https://doi.org/10.1006/jhev.2001.0483
8. Lee-Thorp JA, Van der Merwe NJ, Thackeray JF. The hunters and the hunted revisited. J Hum Evol. 2000;39:565–576. https://doi.org/10.1006/jhev.2000.0436
9. Sponheimer M. Hominins, sedges, and termites: New carbon isotope data from the Sterkfontein valley and Kruger National Park. J Hum Evol. 2005;48:301–331. https://doi.org/10.1016/j.jhevol.2004.11.008
10. Hopley PJ, Maslin MA. Climate-averaging of terrestrial faunas: An example from the Plio-Pleistocene of South Africa. Paleobiology. 2010;36(1):32–50. https://doi.org/10.1666/0094-8373-36.1.32
11. Walker J, Cliff RA, Latham AG. U-Pb isotopic age of the StW 573 hominid from Sterkfontein, South Africa. Science. 2006;314(5805):1592–1594. https://doi.org/10.1126/science.1132916
12. Herries AIR, Shaw J. Palaeomagnetic analysis of the Sterkfontein palaeocave deposits: Implications for the age of the hominin fossils and stone tool industries. J Hum Evol. 2011;60:523–539. https://doi.org/10.1016/j.jhevol.2010.09.001
13. Jones DL, Brock A, McFadden PL. Palaeomagnetic results from the Kromdraai and Sterkfontein hominid sites. S Afr J Sci. 1986;82:160–163.
14. Schmidt VA, Partridge TC. An attempt to establish a magnetostratigraphic framework for sediments deposited in Sterkfontein Cave, South Africa. Eos. 1991;72:44–136.
15. Herries AIR, Curnoe D, Adams JW. A multi-disciplinary seriation of early *Homo* and *Paranthropus* bearing palaeocaves in southern Africa. Quat Int. 2009;202:14–28. https://doi.org/10.1016/j.quaint.2008.05.017
16. Herries AIR, Pickering R, Adams JW, Curnoe D, Warr G, Latham AG, et al. A multi-disciplinary perspective on the age of *Australopithecus* in southern Africa. In: Reed KE, Fleagle JG, Leakey RE, editors. Paleobiology of *Australopithecus*. Vertebrate Paleobiology and Paleoanthropology Series. Dordrecht: Springer; 2013. p. 21–40. https://doi.org/10.1007/978-94-007-5919-0_3
17. Haddad-Martim PM, Hubbe A, Giannini PC, Auler AS, Pilo LB, Hubbe M, et al. Quaternary depositional facies in cave entrances and their relation to landscape evolution: The example of Cuvieri Cave, eastern Brazil. CATENA. 2017;157:372–387. https://doi.org/10.1016/j.catena.2017.05.029
18. Pirson S, Flas D, Abrams G, Bonjean D, Court-Picon M, Di Modica K, et al. Chronostratigraphic context of the Middle to Upper Palaeolithic transition: Recent data from Belgium. Quat Int. 2012;259:78–94. https://doi.org/10.1016/j.quaint.2011.03.035
19. Murszewski A, Edwards TR, Cruden AR, Armstrong B, Boschian G, Herries AIR. Regional geological formation and speleogenesis of the 'Fossil Hominid Sites of South Africa' UNESCO World Heritage Site. Earth-Sci Rev. 2019;188:498–513. https://doi.org/10.1016/j.earscirev.2018.09.016
20. Dreybodt W. Chemical kinetics, speleothem growth and climate. Boreas. 1999;28(3):347–356. https://doi.org/10.1080/030094899422073
21. Palmer AN. Speleogenesis in carbonate rocks. In: Grabrovšek F, editor. Evolution of karst: From prekarst to cessation. Postojna-Ljubljana: Inštitut za raziskovanje krasa, ZRC SAZU; 2002. p. 43–59.
22. Osborne RAL. Cave breakdown by vadose weathering. Int J Speleology. 2002;31:37–53. https://doi.org/10.5038/1827-806X.31.1.3
23. Auler AS, Smart PL, Wang X, Pilo LB, Edwards RL, Cheng H. Cyclic sedimentation in Brazilian caves: Mechanisms and palaeoenvironmental significance. Geomorphology. 2009;106(1–2):142–153. https://doi.org/10.1016/j.geomorph.2008.09.020