South African Journal *of* Science

volume 117
number 1/2

High lightning risk for rural communities in SA

Predicting take-up of home loans using tree-based ensemble models

Monitoring and conservation of terrestrial and marine ecosystems in SA

Hominin lower limb bones from Sterkfontein Caves

HIV self-testing is user-friendly and accurate

ASSAf
ACADEMY OF SCIENCE OF SOUTH AFRICA

Volume 117
Number 1/2
January/February 2021

South African Journal *of* Science

eISSN: 1996-7489

Leader

Book Reviews

Scientific Correspondence

Commentary

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

Design and layout
SUN MeDIA Bloemfontein
T: 051 444 2552
E: publish@sunbloem.co.za

Correspondence and enquiries
sajs@assaf.org.za

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

Profile

Review Article

Research Article

Research Letter

Cover caption
As we enter the UN Decade of Ecosystem Restoration (2021–2031) with the goal "to prevent, halt and reverse the degradation of ecosystems on every continent and in every ocean", several articles in this issue report on the need for or efforts towards monitoring and conservation of terrestrial and marine ecosystems in South Africa (image: ©Nastco).

Celebrating multidisciplinarity

This first issue of the *South African Journal of Science* in 2021 contains a wealth of multidisciplinary articles that are interesting to read while also broadening and deepening topics that are significant to the South African research agenda. While special issues highlight particular topics that consolidate research in a particular field, the Journal's strength lies in its multidisciplinary breadth. We also value the fact that our authors are varied – some are well established in their fields, while others are new to academic publishing and give us the privilege of publishing their first article.

This issue contains a rich variety of Commentaries, which provide perspectives on the link between research and policy, on artificial intelligence and bioinformatics, ecological and social research in the Karoo and other regions, and the impact of COVID-19 on child development. We are also delighted to publish a profile of Professor David Mason of the University of the Witwatersrand, one of South Africa's leading computer scientists and mathematicians.

However, it is the Review and Research Articles that identify this journal issue as remarkable, not only for its content, but for its flagging of many of the Sustainable Development Goals that are crucial to the future well-being of southern Africa and Africa. The article on transdisciplinary research in Namibia by Robert Luetkemeier and colleagues provides direction for others in this complex field that crosses and integrates disciplinary boundaries and holds both promise and difficult challenges. Research Letters by Robin Fisher, Andrew Skowno and co-authors, and Loewan Erasmus et al., speak to environmental research in the marine and savanna environments but also deal with habitat changes in our ecological systems more generally. The Research Article by Tionhonkélé Soro and co-authors identifies the fire hotspots in Côte d'Ivoire, an African country on which more research is always welcome. The high-altitude work of Maqsooda Mahomed, of Dakalo Mashao, and of Moshe Mosotho and their co-authors is a reminder of how much more there is to be discovered about the characteristics of the Earth's atmosphere and how best we may sustain it. The rich and always-fascinating fossil heritage of the sub-continent appears in the hominin work of Travis Pickering and the small mammal work of Thalassa Matthews and their collaborators. Although, for obvious reasons, COVID-19 has gained the most medical attention over the past year, combating and controlling HIV remains the country's other great medical challenge. Mohammed Majam et al. present comparative work on HIV self-test devices and mark progress in this regard. What benefits there may be for higher education through targeted and more generous funding has been analysed by Temwa Moyo and his colleague, and Tanja Verster and her team have been able to investigate important aspects of the home loan industry through her case study.

It is also worth noting the enormous variety of collaborations in all these contributions and to celebrate this networking. Collaboration is evident not only among universities in South Africa, as might be expected – Wits, UCT, UP, UJ, UWC, Rhodes, Free State, North-West, UKZN, Stellenbosch, Nelson Mandela, Mpumalanga, Unisa, Fort Hare – but also in other parts of Africa – Namibia, Nigeria, Zimbabwe and Côte d'Ivoire – and even further abroad. The geographical range of institutional cooperation is impressive: Madison Wisconsin, Colorado, Clemson South Carolina, Birmingham Alabama, Princeton and Minnesota in the USA, Budapest in Hungary, Belfast, Cornwall and Lancaster in the UK, Leuven in Belgium, Bergen in Norway, and Frankfurt in Germany. Moreover, it is also exciting to see collaborations with state structures, such as the South African National Space Agency, the South African Environmental Observation Network, the South African National Biodiversity Institute, Ezemvelo KZN Wildlife and Iziko Museums of South Africa. There is also collaboration with the private sector: ABSA and FNB with Verster et al., and Halteres Associates in San Francisco with Majam et al.

The importance and value of a multidisciplinary journal is not only in that it publishes a diversity of research, but also in that it reaches audiences that, in the ordinary course of their specialised reading, might not be exposed to unfamiliar research areas and from which they can learn much. Journals like the *South African Journal of Science* play a role in breaking down the silos of research and encourage a broad distribution of knowledge to many cohorts of readers who are experts in their own fields, but avid for wider knowledge.

That we are able to present this cornucopia of research is due to the efforts of our team of Associate Editors who shepherd manuscripts through the process of peer review, passing on the comments of peer reviewers in helpful ways and suggesting improvements from their own knowledge and research. This year has not been easy for them as they have grappled with online teaching, a disrupted research agenda of their own, and general scholarly and personal uncertainty. We are also grateful to all our peer reviewers who have taken time and trouble to assess manuscripts for their quality and who are often willing to look at work repeatedly in order to ensure appropriate improvement before publication. During 2020 and the COVID-19 pandemic, many reviewers were even more over-stretched than usual, carrying the burden of teaching online and the uncertainty of the trajectory of the academic enterprise. We therefore doubly appreciate their expertise, willingly provided as volunteers at a difficult time. While our authors and their institutions benefit financially and reputationally from publication in accredited, peer-reviewed, and well-cited journals, those editors and peer reviewers who make this possible do not.

Cumulatively, these contributions and connections attest to the vigour, diversity, and quality of scholarship and academic excellence in our part of the world and we very much hope that you will enjoy reading this issue of the *South African Journal of Science.*

As I end my term of office at the helm of the *South African Journal of Science*, it is a great pleasure to introduce the incoming Editor-in-Chief, Professor Leslie Swartz, a graduate of the University of Cape Town and currently Professor of Psychology at Stellenbosch University. Widely published, internationally well known, and the recipient of many awards, Leslie has been involved in journal publishing in many capacities for decades. He is the Founding Editor and former Editor-in-Chief of the *African Journal of Disability* and the majority of his research lies in the fields of mental health, disabilities studies, and access to health care for vulnerable groups in Africa. We wish him a long and rewarding association with the *South African Journal of Science.*

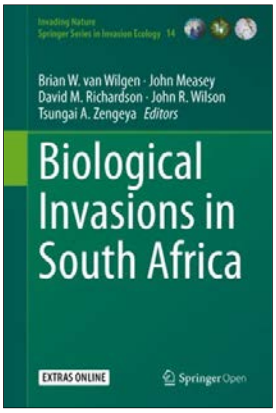
HOW TO CITE:

Carruthers J. Celebrating multidisciplinarity. S Afr J Sci. 2021;117(1/2), Art. #9391. https://doi.org/10.17159/sajs.2021/9391

Invasion science in South Africa: The definitive collection

Check for updates

BOOK TITLE:
Biological invasions in South Africa



EDITORS:
Brian W. van Wilgen, John Measey, David M. Richardson, John R. Wilson, Tsungai A. Zengeya

ISBN:
9783030323943 (ebook, 975 pp)
9783030323936 (hardcover, 975 pp)
9783030323967 (softcover, 975 pp)

PUBLISHER:
Springer, Cham; ZAR950; free ebook

PUBLISHED:
2020

REVIEWER:
Philip E. Hulme[1]

AFFILIATION:
[1]Bio-Protection Research Centre, Lincoln University, Lincoln, New Zealand

EMAIL:
philip.hulme@lincoln.ac.nz

HOW TO CITE:
Hulme PE. Invasion science in South Africa: The definitive collection. S Afr J Sci. 2021;117(1/2), Art. #8732. https://doi.org/10.17159/sajs.2021/8732

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

PUBLISHED:
29 January 2021

Biological invasions by alien pests, weeds and pathogens are a global phenomenon, with increasing impacts on the environment, economy and human health. Traditionally, the poster children of the impacts of biological invasions have been oceanic islands such as Hawai'i and New Zealand, with continental areas being viewed as less prone to the ravages of invasive alien species. Yet this perspective is rapidly changing as continental areas begin to address the threat of biological invasions. For over a decade, the DSI-NRF Centre of Excellence for Invasion Biology (CIB) in South Africa has played a leading role in this changing perspective through outstanding research to reduce the rates and biodiversity impacts of biological invasions by furthering scientific understanding and predictive capability. Now, the major advances in current understanding of biological invasions delivered by the many researchers affiliated to the CIB have been captured in an encyclopaedic volume entitled *Biological Invasions in South Africa*.

The volume sets a new standard in the dissemination of information on biological invasions, not only through its comprehensiveness but also through its philosophy of open access publication, ensuring a global readership irrespective of income or library subscriptions. In the four months after publication in March 2020, the individual chapters were downloaded more than 6000 times each, which reflects the widespread global interest in this topic as well as the considerable outreach achieved by the authors.

The book covers the historical context of biological invasions in South Africa; overviews the invasive taxa and invaded ecosystems; and examines the drivers of invasions, their impacts and subsequent management before exploring new insights from the study of biological invasions in South Africa and tracing out what the future might hold for the country. Many of the invasive species covered in this book are also problems in other parts of the world (e.g. red swamp crayfish, German wasp, pine trees) such that the experiences and progress in South Africa provide valuable lessons to a global audience. South Africa occupies only 2% of the world's land area but it is one of the most biologically diverse countries globally. Thus the chapters play out across this diverse cultural and environmental landscape, bringing to life the values threatened by biological invasions and the urgency of action. The chapters are lavishly illustrated with colour maps and photographs, ensuring that this resource will be valuable both to seasoned practitioners or academic researchers as well as to students just embarking on their careers. Many of the chapters include additional supplementary information that is only available online and which includes a wide range of different material ranging from additional case studies through to promulgated legislation and even detailed data sets.

The coverage is broad, encompassing marine, freshwater and terrestrial ecosystems with reference to plant, insect and vertebrate invaders. However, although the impacts of alien pathogens on vertebrates are described, there is no equivalent chapter for plants, suggesting that the impacts of plant pathogens on South African biodiversity might be underestimated. For example, several pathogens are widespread on introduced crops in South Africa (e.g. myrtle rust and phytophthoras) but little is known of the consequences for native species should these pathogens spill over into natural ecosystems. This is not an oversight of the editors but simply reflects the current state of knowledge – which suggests plant pathology in natural ecosystems would be an area in which greater research investment might be warranted. Indeed, one clear message within this book is that, despite the concerns regarding biological invasions, quantitative assessments of impacts of alien species on biodiversity are relatively few. This is true of many parts of the world and is a significant stumbling block when arguing for a precautionary approach to the introduction of alien species. In the case of South Africa, this lack of information on impacts may also impede the development of a formal biological invasions policy. Yet despite this limited background on impacts, South Africa has led the way in many aspects of invasive species management, especially the use of biological control agents. Some of the most dramatic images from the book are those depicting how landscapes infested with invasive plants changed markedly after the release of biological control agents. The challenges and opportunities for restoring these landscapes following the widespread control of invasive species are discussed in detail. Nevertheless, successful management cannot occur in a vacuum and several chapters touch upon the social dimension of tackling invasive species and especially the situations in which conflicts arise between different groups of stakeholders. There is no easy solution to such conflicts, and once again a key aspect is the need for a clear evidence base that captures the economic, social and environmental benefits of any decisions.

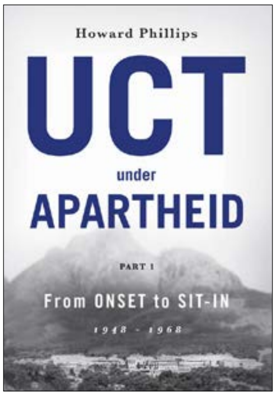
While the undoubted emphasis of the book is on species from other regions that have invaded South Africa, the book points out that this has not been a one-sided exchange and that South Africa has provided the rest of the world with plenty of invasive plant and animal problems as well. My home country of New Zealand seems awash with South African plants but in return a few of our natives, such as our pōhutukawa, have also become invasive in South Africa. These examples illustrate the increasing interconnectedness of world floras and faunas and the need for international collaboration to address these issues. The CIB has been crucial to the internationalisation of research on biological invasions and its contribution to furthering such understanding is provided in the penultimate chapter. A measure of its success is captured in its more than 1700 peer-reviewed publications that include 4237 authors from 110 countries.

The final chapter looks forward to what might be expected in terms of biological invasions in South Africa over the next several decades. The outlook is cautiously positive, indicating that solutions are within reach when science, policy and capacity align. Whatever the future might hold, one thing is for sure: the publication of *Biological Invasions in South Africa* will be seen as a pivotal milestone on the path to achieving these positive outcomes.

The University of Cape Town: Between apartheid and academic freedom

Check for updates

BOOK TITLE:
UCT under apartheid. Part 1: From onset to sit-in, 1948–1968



AUTHOR:
Howard Phillips

ISBN:
9781928232858 (hardcover, 424 pp)

PUBLISHER:
Jacana Media, Johannesburg; ZAR359

PUBLISHED:
2019

REVIEWER:
Tilman Dedering[1]

AFFILIATION:
[1]Department of History, University of South Africa, Pretoria, South Africa

EMAIL:
tdedering@me.com

HOW TO CITE:
Dedering T. The University of Cape Town: Between apartheid and academic freedom. S Afr J Sci. 2021;117(1/2), Art. #8625. https://doi.org/10.17159/sajs.2021/8625

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

PUBLISHED:
29 January 2021

Howard Phillips's history of the University of Cape Town (UCT) from 1948 to 1968 covers a period that is not only remarkable for the gradual consolidation of the apartheid regime but also for important changes in post-World War II academe. The advent of Nationalist Party rule coincided with a phase of increasing professionalisation in a number of scholarly disciplines as well as in the rise of new ideas on the objectives and methodologies of tertiary education. The author has scrutinised the academic, political and social facets of this history by providing a careful examination of the fluctuating trajectories of the University's management, colleges, and departments. Drawing from a wealth of primary, secondary and oral sources, this book also provides plenty of space for the voices of former staff and students. Written in a fluid and elegant style, which efficiently navigates the obstacles of a potentially tedious institutional history, this book takes the reader on a multidisciplinary journey through the South African academic landscape during a period of increasing political oppression.

As must be expected, therefore, the fractious race relations at UCT are never absent from the narrative. The period under discussion saw a vigorous transformation from an institution centred on teaching to one of the most respected research universities in South Africa. This development was accompanied by a remarkable rise in the number of departments from 59 to 105, as well as by the growth of staff numbers from 521 to 893. The development of the different academic disciplines was shaped not only by changing scholarly trends but also by the strengths and weaknesses of more-or-less colourful members of the teaching staff. Occasionally, the author deviates from his studiously diplomatic tone to convey the disappointed reactions from students to pedagogically incompetent teachers or to professors who became stuck in outmoded methodologies.

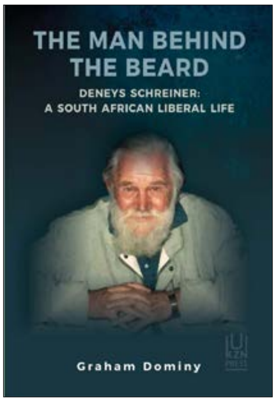
During this period, the student population increased by 73% to 7392 learners (p. 337). The initial hopes of more liberal-minded academics of creating an environment that was more responsive to the educational needs of students from the majority of the South African people, including those classified as coloureds, Indians and Asians, were shattered when the relations between the University management and the government became progressively tense in the 1960s. The author comments on the cherished self-image of UCT as a liberal haven in a sea of racially based injustice with more than a grain of salt. The impressive modernisation of UCT's teaching and research facilities happened without providing the same kinds of services or support to black and coloured students that were afforded to white students. Thus, a sense of alienation was rife among those students whose right to a first-class tertiary education was constantly questioned on racial grounds. While the numbers of coloured and Indian students, who constituted the majority of 'non-white' students, increased from 2.98% in 1948 to 12.4% by 1959, this section of the student population dwindled to a mere 5.5% by 1968 as a result of the intensifying enforced separation between black and white (p. 338).

It has to be acknowledged, as the author does by mentioning individual cases of white defiance, that open confrontation with an increasingly authoritarian and ruthless government was easier to demand than effect. But despite the University's attempts to exploit grey areas and loopholes, its 'timid brand of liberalism', which often glossed over the many instances of segregation in and outside lecture halls, does not lend itself to heroic tales of principled white liberal resistance (p. 253). What often stuck in the minds and memories of African students was that segregation percolated through a liberal veneer to infest most social spaces as they usually characterise campus life. In the wider context of apartheid South Africa, the University's reputation as the most liberal institution (next to Wits) was well deserved, but this also shows what little space the regime left for academic and civic freedom. These political conflicts did not only affect the relations between black and white but also between English and Afrikaans students. Over the years the latter increasingly turned to enrolment at institutions with a pronounced Afrikaner identity because they felt estranged from an English-speaking academic environment that extolled the virtues of a 'foreign' liberalism. Afrikaner nationalist circles did not tire of painting the despised 'Moscow on the Hill' as the hotbed of revolutionary activities, especially when the global wave of student unrest began to lap at the feet of the apartheid colossus in the late 1960s.

This book also provides many insights into gender relations in the academic world during this pre-feminist period. Here the University was very much in line with the Western institutions of higher learning which it always tried to emulate. During the period under discussion, there were only five women among the 65 members of Council, obviously all of them white. The author frequently alludes to a culture of unbridled masculinity that seemed to have been taken for granted by most male and also by female students. For many women, condescending and sexist treatment at the hands of their 'superior' male co-students was a daily experience, especially if they dared to step outside their pre-ordained roles as rag queens and drum majorettes.

For all its engagement with the political and social context, however, this book is a fascinating study of the development of an academic institution during a turbulent period. Experts in the different scholarly fields may find Phillips's succinct descriptions of the emergence of new trends and the concomitant infighting in the different disciplines instructive. It is these struggles against intellectual conformity and stagnation that shape the reputation of a leading university. The author, himself a UCT emeritus professor and a renowned historian, hints at the necessity of a second volume to continue the history of UCT under apartheid. It remains to be hoped that such a volume, bringing up this history to at least the beginning of the democratic South Africa, will indeed follow soon.

BOOK TITLE:
The man behind the beard. Deneys Schreiner: A South African liberal life



AUTHOR:
Graham Dominy

ISBN:
9781869144449 (softcover, 240 pp)

PUBLISHER:
University of KwaZulu-Natal Press, Pietermaritzburg; ZAR295

PUBLISHED:
2020

REVIEWER:
Cas H. Prinsloo[1]

AFFILIATION:
[1]Javett Art Centre, University of Pretoria, Pretoria, South Africa

EMAIL:
cashprinsloo@gmail.com

HOW TO CITE:
Prinsloo CH. In the footsteps of famous family forebears to a respected polymath through moral courage. S Afr J Sci. 2021;117(1/2), Art. #8626. https://doi.org/10.17159/sajs.2021/8626

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

PUBLISHED:
29 January 2021

In the footsteps of famous family forebears to a respected polymath through moral courage

To read *The Man Behind the Beard* is to meet Deneys in person! I picture where and how he lived and worked. I sense the warmth he shared and relive some of his experiences. This is how well Graham Dominy tells the story of Deneys and weaves together a rich flow of events.

My hope is that this review will serve as the appetiser for the main meal and sumptuous dessert served by Dominy: a mental picture of architecture, sights and sounds, south and north, high ground and open spaces, joys and heartaches.

Colin Gardner[1] described Deneys in his obituary as '… a good man, a fine scholar and a clever thinker who made a difference at many levels in the lives of thousands: colleagues, students, the marginalised and the hopeful. All this was framed by his beard.' (quoted on p.199) To this I would add that he was concerned about justice, academic freedom and integrity, and social and educational development. But no, I'm not going to say when the beard started growing -- only that when it became time for Deneys to shave it off, he also used the occasion to play a trick on a close family member!

The book's chronology tells of the development of Deneys's character; details the influences on his life and their outcomes; and provides insight into him as an individual in the context of his forebears, childhood, youth, wartime service, and studies in South Africa, Britain and America. In the process, his family life, his research and academic career, including tertiary institution management, are linked with his political awareness and contributions. Of course, Deneys's life story unfolded in the context of colonialism and its aftermath, still so pertinent today. Deneys was not spared deep family tragedy, and the reality and emotion thereof will resonate with most readers.

Dominy portrays a man whose life was enriched through nature, culture, sport, art, politics, literature, history, economics, science, and open and meaningful relations with family, friends and colleagues irrespective of domain, sphere or hierarchy. He showed courage, intellectual power, moral character and humour, all of which may powerfully and vicariously guide us now and in the future.

Writing a review always poses the risk of spoiling a good read. I therefore do not reveal too many facts and details. It is sufficient to say that *The Man Behind the Beard* is very accessible, and should be of interest to a wide, multidisciplinary readership. George Deneys Lyndall Schreiner walked this earth from 1923 to 2008. The observant reader will immediately note that he lived, fully aware, through a world war and the momentous upheavals and transformations in South Africa. Through all of this he became a geophysicist of standing, a true scientist and academic; developed a natural teaching ability; served a stint as researcher at Wits and as Dean of Science of the University of KwaZulu-Natal (UKZN); married a wonderful life partner, Else – his equal or better, who gave him support of note; fathered and raised together two sons and two daughters; and became Vice-Principal of UKZN, with all the administrative (read committee) work, managerial, academic and other leadership challenges it entailed. His contributions as a liberal (thinker and practitioner), in the best sense of the word, are what stand out in the struggle for democracy. Although he had no formal legal, political or constitutional background, his transformative work at UKZN, and the enormous role he played as founder member and chairperson of the so-called Buthelezi Commission, were most likely the two biggest feathers in his cap. Dominy sensitively, in a nuanced manner, relates how Deneys and his co-cast chose to resist inequality, unfairness and discrimination – some through armed struggle and others through different forums and strategies; some by choice and others by uncomfortable compromise; and some from within the system and others not.

Perhaps a brief word on family lineage is fitting. Olive Schreiner was Deneys's great-aunt. His grandfather and father, respectively, were Prime Minister of the Cape and Judge of the Appellate Division of the Supreme Court of South Africa. His father could easily have become Chief Justice, politics permitting. Deneys and Else's children followed this lineage. One daughter, a member of Umkhonto we Sizwe, served jail time of about four years, complete with all its harrowing experiences, through one of the final trials conducted by the outgoing apartheid regime, before being pardoned. Everyone in the family practised their citizenship with commitment.

The book is a biographical work in the finest tradition of historiography. The content covers a humble, but critical, figure in the history of South Africa who witnessed critical developments that often helped shape pertinent outcomes.

Dominy, former National Archivist of South Africa, made many important contributions on the African continent and abroad in documenting aspects of heritage. The author shows true craftsmanship in tracing facts from a staggering array of sources. These included interviews with family, former colleagues, friends and journalists; many weeks scouring archive collections; and studying the usual range of formal publications. He then, in the second act of wizardry, artfully wove the raw material into a tale; a narration (subtly witty) that holds your attention. The craftmanship is on par with that of the well-known historical novelist, Simon Winchester.

The foreword may reveal more about Deneys and the book, being written by the current South African Minister of Higher Education, Science and Technology, Blade Nzimande. This is no coincidence and should also not be perceived cynically as nifty footwork to endorse the author or publisher. The foreword is borne from the deepfelt, close relationship between Deneys and Blade Nzimande. Schreiner served as mentor to many people and was the guardian of their access to tertiary learning in difficult times. The foreword reflects an appreciation for each other's worth and values, grown during regular Sunday-evening dinners at the Schreiner's homestead in Pietermaritzburg.

The excellent structure and pace of the book are complemented by a categorised bibliography, an index, notes and acknowledgements. In summary, succinct formulation, a natural flow, interesting presentation and the clever use of wit make for a good read!

References

1. Gardner C. Obituary: George Deneys Lyndall Schreiner (1923–2008). Natalia. 2008;38:83–85.

AUTHORS:
Mark Tomlinson[1,2]
Linda Richter[3]
Wiedaad Slemming[4]

AFFILIATIONS:
[1]Institute for Life Course Health Research, Department of Global Health, Stellenbosch University, Cape Town, South Africa
[2]School of Nursing and Midwifery, Queens University, Belfast, United Kingdom
[3]DST-NRF Centre of Excellence in Human Development, University of the Witwatersrand, Johannesburg, South Africa
[4]Division of Community Paediatrics, Department of Paediatrics and Child Health, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa

CORRESPONDENCE TO:
Mark Tomlinson

EMAIL:
markt@sun.ac.za

HOW TO CITE:
Tomlinson M, Richter L, Slemming W. What the science of child and adolescent development contributes to understanding the impacts of COVID-19. S Afr J Sci. 2021;117(1/2), Art. #8876. https://doi.org/10.17159/sajs.2021/8876

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
nurturing care, child development, indirect effects of COVID-19, support networks, facilitating environments

PUBLISHED:
29 January 2021

What the science of child and adolescent development contributes to understanding the impacts of COVID-19

As of 8 September 2020, there were 27 236 916 confirmed cases of COVID-19, including 891 031 deaths reported to the World Health Organization.[1] The direct effects of SARS-CoV-2 (the virus) and COVID-19 (the disease caused by the virus) are infection (of which ~80% of people will have no or mild symptoms), serious illness (~15% requiring hospitalisation and ~5% ventilation) and death (<1%). Of all age groups, children (0–18 years of age) have the lowest risk of the direct effects of the virus and the disease. Understandably, because of this, much of the global focus has been on protecting the elderly and adults with co-morbidities and using country lockdowns to prevent community transmission. What has often been forgotten, however, is that children have the highest risk of all age groups of experiencing the indirect adverse effects of the pandemic and the effects to contain it. As ~20% of deaths occur amongst people over 65 years of age, many children will lose beloved grandparents and older relatives, and many will lose their primary caregiver. Newborn and young children may be separated from their mothers – an experience that can have long-term effects on children's development.[2] Other indirect effects result from actions and the knock-on effects of actions taken to prevent, control and treat the virus. These include societal lockdowns, isolation and quarantine with follow-on negative effects on income and food security, fear and panic, anxiety and depression, altered family and social relations, stigma and, in extreme cases, post-traumatic stress.[3]

The data on the consequences of the pandemic are stark: globally it is predicted that more than a million preventable child deaths will occur[4]; a 'hunger pandemic' is already happening; and tens of millions more children are falling into extreme poverty[5]. Rising rates of malnutrition and stunting are going to massively impact children's cognitive trajectories across the life course.[6] Educationally, the effects on children will be catastrophic, with approximately 1.5 billion children not benefitting from continuous schooling due to rolling lockdowns.[7] We know that for many children, not being in school for a protracted period of time means that they will never return to school[8]; that for millions of children without digital access there will be no 'home schooling'; and it will be the poorer children that will fall even further behind[7]. Most worryingly for disadvantaged families, hundreds of millions of children will not be receiving school meals.

Children of all ages depend on parents and families for health care, nutritious food, protection from harm, opportunities to learn, and love and affection – together comprising *nurturing care*.[9] The youngest children depend on nurturing care to both survive and thrive.[10] The ability of families to provide nurturing care for their children is highly dependent on facilitating environments. Facilitating environments are made up of, amongst others, availability of decent work, housing, health care, social security and laws that support families. As children develop, sources of nurturing care extend to include the wider family network, childcare workers, teachers, community members and, very importantly, friends and peers.[11]

Nurturing care can break down when features of the facilitating environment are perturbed, as has happened during lockdown, for example, and when social relations are disrupted, as occurs with isolation and quarantine. Lack of income, interruption of health services, and poorly implemented social safety nets threaten the health and well-being of children directly. In addition to fear engendered by media and conversations around them, the ability of adults to protect them may be eroded by their own rising anxiety and mental health concerns. People react to stressful situations in different ways. Some may withdraw, becoming cold, distant and emotionally unavailable, which is especially damaging for young children and known to have long-term consequences on their mental and physical health.[12] Others may experience breakdowns in their social and psychological controls, putting them on edge, easily angered and with lowered restraints on lashing out verbally and physically at others, including children. Violence, especially when children are trapped in a space with an aggressive adult, is frightening and especially harmful if recurrent.

A major source of resilience for children is supportive adults beyond the household, such as mentors, extended family and their educators. Lockdowns, isolation, the closure of schools and separation from friends interrupt the usual balancing of adverse and protective experiences that enable children to cope on a day-to-day basis. They also interrupt the rich informal learning that occurs in children's interactions with teachers, coaches, mentors and friends that complement formal school education in imparting knowledge and skills that are critical to personal and social adjustment. Networks of friends are especially important for adolescents, in helping young people to negotiate their growing independence and the demands of family, school and society.[13] Being confined to home with parents whose vigilance and control is heightened by their own fear, is especially frustrating for adolescents, who may themselves withdraw or lash out.[13]

There is no doubt that some children and adolescents will be damaged by the direct and indirect effects of SARS-CoV-2 and COVID-19. The century-old science of child development demonstrates the harmful effects of babies separated from mothers in their first months after birth; of children experiencing uncontrollable fear without the mediating reassurance of confident adults and food secure environments; the long-term consequences on the health and well-being of children and adolescents living with withdrawn or depressed parents, and of the physical, emotional and sexual abuse of children.[2] The damage experienced in childhood, especially among young children who have not yet developed some independent mechanisms of emotional regulation, has been shown to manifest throughout life in poor school achievement, lack of healthy and satisfying relationships, unstable mental health, lower levels of work productivity and less social stability.[6]

On the flip side, children who experience the nurturing care of family and friends and who live in environments that meet their basic needs, have enormous reservoirs of resilience.[14] Preserving child and family well-being will ensure that children remain developmentally on track and will help to offset the effects of the ongoing, varied and increasing adversity that many are faced with at this time.[15] Children's coping mechanisms are strengthened by restoring daily routines, by engaging them actively in solving and resolving difficulties and by the support of group activities with their peers.[16] Children can be powerful health promoters, as demonstrated by child-to-child campaigns, and eager adopters of rules and guidelines to protect themselves and others. Thus, they can be active agents in supporting public health approaches to dealing with the pandemic.

The foundations of human well-being are laid down in childhood, and it is no hyperbole to state that the future well-being of our societies depends on the enabling environments we create and nurturing care we provide for our children today and across their life course. Neglecting the profound implications of this pandemic on children – both in the present as well as in the immediate post-COVID-19 world – will be a key determinant of whether our societies thrive in the future. We have a window of opportunity to ensure that just because children and adolescents carry the lowest risk of infection, we do not consign them to a multitude of other risks across their lifetime.

References

1. World Health Organization. WHO Coronavirus disease (COVID-19) dashboard 2020 [webpage on the Internet]. No date [cited 2020 Sep 09]. Available from: https://covid19.who.int/?gclid=Cj0KCQjw-uH6BRDQARIsAI3I-UdHtRYCziwB0foTkKqV-UdXJNQ4eVB8Gh63tZd8mMSQ5BGNeDi4kbAaAttGEALw_wcB.
2. Shonkoff JP, Richter L, Van der Gaag J, Bhutta ZA. An integrated scientific framework for child survival and early childhood development. Pediatrics. 2012;129(2):460–472. https://doi.org/10.1542/peds.2011-0366
3. Holmes EA, O'Connor RC, Perry VH, Tracey I, Wessely S, Arseneault L, et al. Multidisciplinary research priorities for the COVID-19 pandemic: A call for action for mental health science. Lancet Psychiatry. 2020;7(6):547–560.
4. Roberton T, Carter ED, Chou VB, Stegmuller AR, Jackson BD, Tam Y, et al. Early estimates of the indirect effects of the COVID-19 pandemic on maternal and child mortality in low-income and middle-income countries: A modelling study. Lancet Glob Health. 2020;8(7):901–908. https://doi.org/10.1016/s2214-109x(20)30229-1
5. Fore HH, Dongyu Q, Beasley DM, Ghebreyesus TA. Child malnutrition and COVID-19: The time to act is now. Lancet. 2020;396(10250):517–518. https://doi.org/10.1016/s0140-6736(20)31648-2
6. Black MM, Walker SP, Fernald LCH, Andersen CT, DiGirolamo AM, Lu C, et al. Early childhood development coming of age: Science through the life course. Lancet. 2017;389(10064):77–90. https://doi.org/10.1016/s0140-6736(16)31389-7
7. United Nations Children's Fund. Keeping the world's children learning through COVID-19. New York: UNICEF; 2020.
8. Save the Children. Save our Education. London: Save the Children; 2020.
9. World Health Organization. Nurturing care for early childhood development: A framework for helping children survive and thrive to transform health and human potential. Geneva: World Health Organization; 2018.
10. United Nations. The Global Strategy for Women's, Children's and Adolescents' Health (2016-2030) New York: United Nations; 2015.
11. Richter LM, Daelmans B, Lombardi J, Heymann J, Boo FL, Behrman JR, et al. Investing in the foundation of sustainable development: Pathways to scale up for early childhood development. Lancet. 2017;389(10064):103–118. https://doi.org/10.1016/s0140-6736(16)31698-1
12. Tomlinson M, Cooper P, Murray L. The mother-infant relationship and infant attachment in a South African peri-urban settlement. Child Dev. 2005;76(5):1044–1054. https://doi.org/10.1111/j.1467-8624.2005.00896.x
13. Patton GC, Sawyer SM, Santelli JS, Ross DA, Afifi R, Allen NB, et al. Our future: A Lancet commission on adolescent health and wellbeing. Lancet. 2016;387(10036):2423–2478.
14. Luthar SS, Cicchetti D, Becker B. The construct of resilience: a critical evaluation and guidelines for future work. Child Dev. 2000;71(3):543–562.
15. Tomlinson M, Ross DA, Bahl R, Rollins N, Daelmans B, Simon J, et al. What will it take for children and adolescents to thrive? The Global Strategy for Women's, Children's, and Adolescents' Health. Lancet Child Adolesc Health. 2019;3(4):208–209. https://doi.org/10.1016/s2352-4642(19)30004-5
16. Clark H, Coll-Seck AM, Banerjee A, Peterson S, Dalglish SL, Ameratunga S, et al. A future for the world's children? A WHO-UNICEF-Lancet Commission. Lancet. 2020;395(10224):605–658.

AUTHOR:
Chenjerayi Kashangura[1]

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

CORRESPONDENCE TO:
Chenjerayi Kashangura

EMAIL:
ckashangura@kutsaga.co.zw

HOW TO CITE:
Kashangura C. Artificial intelligence enhanced molecular databases can enable improved user-friendly bioinformatics and pave the way for novel applications. S Afr J Sci. 2021;117(1/2), Art. #8151. https://doi.org/10.17159/sajs.2021/8151

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
deep learning, deep reasoning, algorithm

PUBLISHED:
29 January 2021

Artificial intelligence enhanced molecular databases can enable improved user-friendly bioinformatics and pave the way for novel applications

Molecular databases have enabled scientists across the globe to collaborate and contribute to the growth of the databases. The current form of the databases involves researcher input which is acted upon by algorithms developed by bioinformaticians leading to outputs for researchers. Experimental data analysis using the molecular databases normally results in a reduction in cost and time for in vitro experiments preceded by in silico stages. Molecular biology technologies are applied in multiple disciplines, generating enormous amounts of data every day, which, when deposited, requires professional staff to annotate, verify submissions and generally maintain the database. The rapid rise of artificial intelligence (AI) can be used to enhance molecular databases through incorporation of deep learning and deep reasoning to enable the molecular databases to partially self-maintain, bringing novel applications and the potential for an improved user-friendly interface for researchers who are not trained in bioinformatics to generate data that require bioinformatics-related analysis.

Bioinformatics has been around since the 1960s, whilst online molecular databases that handle data generated by disciplines in the life sciences have existed since the 1990s[1], with researchers contributing by submitting data generated through experiments conducted in vitro and in vivo or by utilising the molecular databases for in silico analysis. The backbone of this analysis is bioinformatics presented in various algorithms tailor-made for different molecular data, such as DNA (genomics), RNA (transcriptomics) and protein (proteomics), to produce particular outcomes. This analysis relies on the researcher interrogating the database through the algorithms.

A plethora of databases such as GenBank has been published in the *Nucleic Acids Research* Database issues for the past 26 years, highlighting a range of different molecular data and a synchronous range of bioinformatics capabilities within the databases.[2-4] The tools to use on data contained in the molecular databases is determined by researchers. Often a researcher may opt to use the bioinformatics tools they are well acquainted with and opt not to use other tools which might require arduous training. However, the data submitted to providers of molecular databases require skilled professionals to update and secure systems and verify the accuracy of the data submitted[1] and funds are required for such staff.

The growing availability of AI brings the possibility of self-learning, self-reasoning and self-improving molecular databases that can be considered to be 'next-generation enhanced molecular databases' which can assist in lowering the cost of maintenance, ultimately reducing databases becoming defunct and introducing novel applications and a new in-depth analysis of molecular data. These next-generation databases can enable bioinformatics to become more user-friendly to non-bioinformatics trained researchers who, owing to the multi- and interdisciplinary nature of molecular life sciences data, sometimes face a daunting task in analysing data.

This next generation may take the form of AI led or incorporated databases. These can have various forms of deep learning, deep reasoning and reading algorithms that enable the databases to learn experientially. These databases will not be static but will be capable of increasing the database routine tasks that can be accomplished through experiential learning so that activities such as curation, annotation and archiving can be partly accomplished, with human verification required, together with the introduction of new potential applications.

Molecular databases enhanced with AI may be useful for a new level of deep analysis that involves the AI selecting the parameters to be used through experiential learning in areas that include three-dimensional modelling[5], binding predictions and domain calling[6], epigenome, especially differential DNA methylation patterns[7], deep analysis of multi-omic data[8,9], sequence-based taxonomy[3], precision medicine and drug discovery[10,11]. The experiential learning capability suggests that the database can, for example, interrogate a submitted set of molecular data and determine the nature of the data and carry out basic analysis to assign identity, annotation and generation of output from the AI selected parameters, like a phylogenetic tree or identified potential antimicrobial agents. A user interface can then pop up with suggestions for further data analysis, for example, if the AI has identified a potential therapeutic agent against existing or emerging pathogens, then prediction AI simulations[12] are selected and run using deep learning algorithms that would run in silico trained parameter algorithms for predicting pathogen 'growth inhibition', or 'neutralise pathogen receptor access', or 'prevent multiplication'. This will greatly assist those researchers who are not bioinformatics trained to generate molecular life sciences data that require bioinformatics analysis. These next-generation databases can potentially provide improved user-friendly interaction with bioinformatics by providing single-click buttons for running particular bioinformatics tools whilst the parameters are selected by the AI after analysis of the submitted data set.

The current challenge of big data analysis may be ameliorated by development of AI that performs a comparative analysis of recently submitted big data sets against existing similar data sets and possibly suggests areas that need modification in terms of analysis for bioinformaticians to develop.

Competing interests

I declare that there are no competing interests.

References

1. Imker HJ. 25 Years of molecular biology databases: A study of proliferation, impact and maintenance. Front Res Metr Anal. 2018;3, Art. #18, 13 pages. https://doi.org/10.3389/frma.2018.00018

2. Johnson G, Wu TT. Kabat Database and its applications: 30 Years after the first variability plot. Nucleic Acids Res. 2000;28(1):214–218. https://doi.org/10.1093/nar/28.1.214

3. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, et al. GenBank. Nucleic Acids Res. 2013;41(D1):D36–D42. https://doi.org/10.1093/nar/gks1195

4. Rigden DJ, Fernández XM. The 26th annual Nucleic Acids Research database issue and molecular biology database collection. Nucleic Acids Res. 2019;47(D1):D1–D7. https://doi.org/10.1093/nar/gky1267

5. Ay F, Noble WS. Analysis methods for studying the 3D architecture of the genome. Genome Biol. 2015;16(1), Art. #183, 15 pages. https://doi.org/10.1186/s13059-015-0745-7

6. Joyce AP, Zhang C, Bradley P, Havranek, JJ. Structure-based modelling of protein: DNA specificity. Brief Funct Genomics. 2014;14(1):39–49. https://doi.org/10.1093/bfgp/elu044

7. Huh I, Wu X, Park T, Yi SV. Detecting differential DNA methylation from sequencing of bisulfite converted DNA of diverse species. Brief Bioinform. 2019;20(1):33–46. https://doi.org/10.1093/bib/bbx077

8. Sangalaringam A, Ullah AZD, Marzek J, Gadaleta E, Nagano A, Ross-Adams H, et al. 'Multi-omic' data analysis using O-miner. Brief Bioinform. 2019;20(1):130–143. https://doi.org/10.1093/bib/bbx080

9. Spies D, Renz PF, Beyer TA, Ciaudo C. Comparative analysis of differential gene expression tools for RNA sequencing time course data. Brief Bioinform. 2019;20(1):288–289. https://doi.org/10.1093/bib/bbx115

10. Stokes JM, Yang K, Swanson K, Jin W, Cubillos-Ruiz A, Donghia NM, et al. A deep learning approach to antibiotic discovery. Cell. 2020;180(4):688–702. https://doi.org/10.1016/j.cell.2020.01.021

11. Hoffman M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, et al. SARS-Cov-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically-proven protease inhibitor. Cell. 2020;181(2):271–280. https://doi.org/10.1016/j.cell.2020.02.052

12. Rodriguez AC. Simulation of genes and genomes forward in time. Curr Genomics. 2010;11(1):58–61. https://doi.org/10.2174/138920210790218007

Check for updates

AUTHORS:
Lain E. Pardo[1,6]
Sara Bombaci[2]
Sarah E. Huebner[3]
Michael J. Somers[4,5]
Herve Fritz[1,6]
Colleen Downs[7]
Abby Guthmann[3]
Robyn S. Hetem[8]
Mark Keith[4]
Aliza le Roux[9,10]
Nokubonga Mgqatsa[11]
Craig Packer[3]
Meredith S. Palmer[12]
Daniel M. Parker[9,13]
Mike Peel[7,14,15]
Rob Slotow[7]
W. Maartin Strauss[16]
Lourens Swanepoel[17,18]
Craig Tambling[19]
Nairobi Tsie[4]
Mika Vermeulen[1,6]
Marco Willi[20]
David S. Jachowski[7,21]
Jan A. Venter[1,6]

AFFILIATIONS:
[1]School of Natural Resource Management, Nelson Mandela University, George, South Africa
[2]Department of Fish, Wildlife, and Conservation Biology, Colorado State University, Fort Collins, Colorado, USA
[3]College of Biological Sciences, University of Minnesota, St. Paul, Minnesota, USA
[4]Eugène Marais Chair of Wildlife Management, Mammal Research Institute, Department of Zoology and Entomology, University of Pretoria, Pretoria, South Africa
[5]Centre for Invasion Biology, University of Pretoria, Pretoria, South Africa
[6]REHABS, International Research Laboratory, French National Centre for Scientific Research (CNRS) / University of Lyon 1 / Nelson Mandela University, George, South Africa
[7]School of Life Sciences, University of KwaZulu-Natal, Durban, South Africa
[8]School of Animal, Plant and Environmental Sciences, University of the Witwatersrand, Johannesburg, South Africa
[9]Department of Zoology and Entomology, University of the Free State, Phuthaditjhaba, South Africa
[10]Afromontane Research Unit, University of the Free State, Phuthaditjhaba, South Africa
[11]Wildlife and Reserve Management Research Group, Department of Zoology and Entomology, Rhodes University, Makhanda, South Africa
[12]Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey, USA
[13]School of Biology and Environmental Sciences, University of Mpumalanga, Mbombela, South Africa
[14]Agricultural Research Council, Animal Production Institute, Rangeland Ecology, Pretoria, South Africa

Snapshot Safari: A large-scale collaborative to monitor Africa's remarkable biodiversity

Nature is experiencing degradation and extinction rates never recorded before in the history of Earth.[1,2] Consequently, continuous large-scale monitoring programmes are critical, not only to provide insights into population trends but also to aid in understanding factors associated with altering population dynamics at various temporal and spatial scales.[3] Continuous monitoring is important not only for tracking rare or threatened species but also to detect the increase of potentially invasive species[4], and the trends in the populations of common species, which in some regions are declining even more rapidly than are rare species[2].

The combination of citizen science and cutting-edge technologies has improved monitoring programmes.[5] In this regard, camera traps have become a popular tool to engage with society while obtaining accurate scientific data.[3] The importance of advances in technological monitoring has even been highlighted by the United Nations Environment Programme (UNEP) through the proposed 'Digital Ecosystem framework', a complex distributed network or interconnected socio-technological system.[6]

Monitoring species and ecosystems is critical to Africa – a highly biodiverse continent with numerous mammal species threatened by human activities such as poaching, overhunting, and climate and land-use change.[7] Over half the terrestrial mammals in Africa have experienced range contractions of as much as 80% on average, including predator species such as lions (*Panthera leo*) and large ungulates.[2] In sub-Saharan Africa, human impacts are projected to increase, and trigger an increased extinction risk.[7] However, information on the conservation status of many species is limited, and many areas in Africa lack the baseline biodiversity data necessary to assess the outcomes of existing conservation programmes.[5] Further, the lack of standardised methods to assess biodiversity patterns limits our ability to detect and respond to changes in mammal populations caused by environmental and anthropogenic factors.[8]

In attempting to address some of the above challenges, we have formed the Snapshot Safari Network (www.snapshotsafari.org) – a large-scale international camera trap network to study and monitor the diversity and ecological dynamics of southern and eastern African mammals. Snapshot Safari (hereafter Snapshot) is one of the largest camera trap networks in the world. It began in 2010 with a single camera trap grid in Serengeti National Park, Tanzania[9], and the model and protocols have since been expanded in Tanzania as well as into five other countries: Botswana, Kenya, Mozambique, South Africa, and Zimbabwe (Figure 1). Participating locations represent a wide variety of habitats, wildlife communities, management types and protected area sizes. Here, we introduce this multidisciplinary initiative which combines citizen science and advanced machine learning techniques for the analysis of millions of animal photographs. We also introduce a set of high priority research questions emanating from expert consultation in 2019.

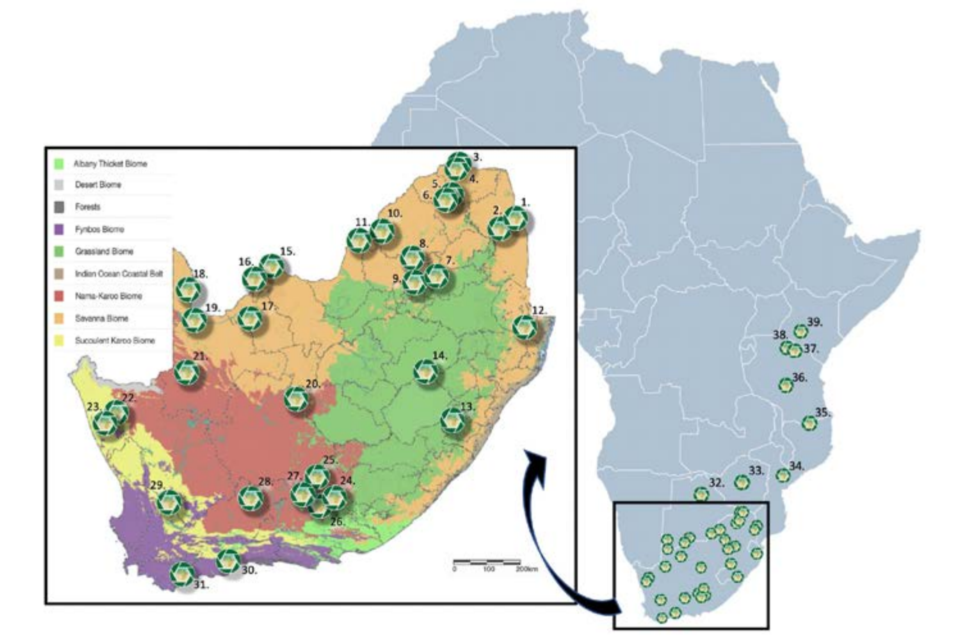


Figure 1. Snapshot Safari grid sites: **1.** Kruger National Park **2.** Associated Private Nature Reserves **3.** Mapungubwe Nature Reserve **4.** Venetia Nature Reserve **5.** Langjan Nature Reserve **6.** Blouberg Nature Reserve **7.** Telperion Private Nature Reserve **8.** Pilanesberg National Park **9.** Rietvlei Nature Reserve **10.** Atherstone Nature Reserve **11.** Madikwe Game Reserve **12.** Somkhanda Game Reserve **13.** Ntsikeni Wildlife Reserve **14.** Golden Gate Highlands National Park **15.** Khamab Kalahari Game Reserve **16.** Molopo Nature Reserve **17.** Tswalu Kalahari Reserve **18.** Kgalagadi Transfrontier Park **19.** Khomani San Common Property Association **20.** Mokala National Park **21.** Augrabies Falls National Park **22.** Goegap Nature Reserve **23.** Namaqua National Park **24.** Mountain Zebra National Park **25.** Bergplaas Nature Reserve **26.** Samara Private Nature Reserve & De Nuk Farm **27.** Camdeboo National Park **28.** Karoo National Park **29.** Tankwa Karoo National Park **30.** Gondwana Game Reserve **31.** DeHoop Nature Reserve and Overberg **32.** Magkadigkadi Pans National Park **33.** Debshan Ranch **34.** Gorongosa National Park **35.** Niassa National Park (Mariri) **36.** Ruaha National Park **37.** Serengeti National Park **38.** Grumeti Game Reserve **39.** Enonkishu Conservancy

Figure 1: Current Snapshot Safari study locations in southern Africa.

[15]Applied Behavioural Ecology and Ecosystems Research Unit, University of South Africa, Johannesburg, South Africa
[16]Department of Environmental Sciences, University of South Africa, Johannesburg, South Africa
[17]Department of Zoology, University of Venda, Thohoyandou, South Africa
[18]African Institute for Conservation Ecology, Makhado, South Africa
[19]Department of Zoology and Entomology, University of Fort Hare, Alice, South Africa
[20]School of Physics and Astronomy, University of Minnesota, Minneapolis, Minnesota, USA
[21]Department of Forestry and Environmental Conservation, Clemson University, Clemson, South Carolina, USA

CORRESPONDENCE TO:
Sarah Huebner

EMAIL:
huebn090@umn.edu

HOW TO CITE:
Pardo LE, Bombaci S, Huebner SE, Somers MJ, Fritz H, Guthmann A, et al. Snapshot Safari: A large-scale collaborative to monitor Africa's remarkable biodiversity. S Afr J Sci. 2021;117(1/2), Art. #8134. https://doi.org/10.17159/sajs.2021/8134

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
citizen science, camera trap, conservation, machine learning, mammals

PUBLISHED:
29 January 2021

Standardised methods

Our camera trap design consists of regular grids of 5 km^2 per location (grid sites). Each grid has 8 to 245 camera traps depending on the objectives and area to cover. Cameras are secured with steel cases and fixed at ~50 cm height to detect medium- to large-bodied mammals. Each camera is programmed to take a series of three images within 1–5 seconds of each other (a 'capture event') when passive infrared sensors are triggered by motion or heat during the day and one image at night.[9] Most grids have operated continuously since 2018and are intended to run for a decade or longer. Data collected are forwarded to the University of Minnesota Lion Center for curation and management of the citizen science component. Metadata on camera placements and habitat characteristics are also collected in a standardised manner to facilitate cross-site comparisons.

Camera traps generate large volumes of photographs which makes the classification of species a time-consuming task. To facilitate efficient image processing, we combine citizen science efforts with advanced machine learning techniques.[10] Snapshot partners with the citizen science platform www.zooniverse.org, on which volunteers identify species and annotate other information such as the number of individuals or behaviour. They can also interact directly with researchers on the talk boards or via social media. Each of our 24 current projects has its own webpage within the Snapshot organisation. More than 150 000 volunteers worldwide have classified over 9 million photographs since the relaunch of Snapshot as a network in February 2018. These responses exhibit 97% accuracy, confirming the reliability of citizen science in rapidly processing large volumes of data.[11]

Machine learning

The Snapshot network also incorporates machine learning algorithms prior to uploading data to Zooniverse to decrease the number of citizen scientists required to view each capture event.[10] Two image classifiers (convolutional neural networks) are employed; one to identify empty images and one to predict the species, counts, and behaviours in images containing wildlife. The algorithms' predictions and confidence levels are uploaded to Zooniverse with the image manifest and used to dynamically determine the level of agreement and number of volunteers required to confirm the algorithms' label (see Figure 2). The millions of images generated annually by the Serengeti grid have been utilised as training data for many deep learning algorithms developed to identify African mammal species automatically.[10,12,13]

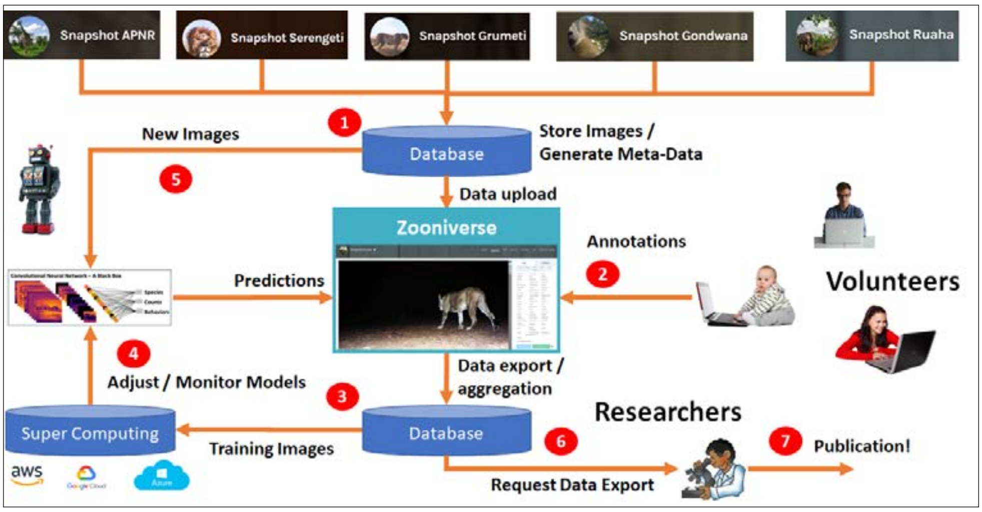


Figure 2: Snapshot Safari data processing workflow. **(1)** Data from each location is checked for errors (bad flash, date-time reset) and run through machine learning algorithms during the pre-processing phase. **(2)** Data and machine learning predictions are uploaded to Zooniverse for classification. Every image is classified by multiple volunteers, and their votes are aggregated to assign labels of species, count, and behaviours. **(3–5)** Citizen scientists' labelled annotations are used to iteratively train and refine the machine learning algorithm. **(6–7)** Data are returned to researchers and reserve management for publications and conservation assessments.

The first stage of classification occurs on the Zooniverse in a binary workflow, in which volunteers evaluate whether animals can be seen in the images. If two people agree with the algorithm that an image is empty, it is removed from the data set. If one person disagrees with the computer, the image remains in circulation to accumulate three more human votes, at which point majority consensus is accepted. Images marked as containing wildlife are moved to the second stage of classification, in which the species, count, and behaviours are annotated. At this stage, dynamic rules within the Zooniverse infrastructure are used to retire captures based on the species, agreement of the volunteers with the computer and one another, and the algorithm's confidence in its own prediction.

As a result of quickly removing blank images and those of humans and common species, citizen scientists are presented with more images of rare and cryptic species, improving the volunteer experience[14] and providing valuable data to refine the machine learning algorithm's capabilities. This approach has improved efficiency by 43% thus far on the Serengeti project.[10] Citizen science and machine learning stages are applied iteratively to constantly improve performance and efficiency. The labelled images from all projects retrain the existing model to improve the machine's predictive capability in a variety of habitats, and the refined model is run on new data sets as they are collected from the field (Figure 2).

Snapshot also provides useful data for computer scientists who wish to build artificial intelligence algorithms to identify wildlife species automatically without human assistance. Data from six South African sites and the Serengeti are publicly available for download on the Microsoft-hosted site Labelled Image Library of Alexandria – Biology & Conservation (LILA-BC) at www.lila.science.com. More data are added to the labelled image repository as datasets are finalised. In all cases, images of humans and rhinoceros are withheld from the repository for privacy and poaching concerns, respectively. Metadata on habitat and environmental characteristics are available to researchers outside of the network upon request. This ensures that local researchers and reserve management are aware of how the data are being used and can contribute and collaborate.

Outputs and general results

Snapshot is documenting the presence, distribution, diversity, and ecology of at least 85 medium- to large-sized African mammals. The variety of Snapshot locations provides numerous opportunities to answer questions in wildlife ecology and conservation, test ecological hypotheses and analytical methods, and measure the impacts of anthropogenic disturbances across multiple spatiotemporal scales. For example, the data produced by the camera traps running continuously in the Serengeti provided the basis for papers on spatiotemporal partitioning[15-17], behavioural interactions[18-20], and advancing modelling techniques used with the camera trap data[11,19,21].

Scientists visit sites regularly to maintain and manage the camera trap grids and also to improve community participation. For example, in Ruaha National Park (Tanzania), researchers trained community members to deploy and maintain the grid[22], thus creating jobs and infrastructure. In South Africa, we are building strong relationships with local communities, such as the Khomani San in the Kalahari and managers of private reserves, who will use the camera trap information to improve their wildlife management practices.

Snapshot Safari – South Africa

Like other African countries, South Africa lacks adequate conservation data for many species, and about 17% of mammal species are threatened with extinction.[8] However, South Africa has a long history of conservation intervention and is, therefore, a testing ground for different hypotheses. In South Africa, we have surveyed 31 locations (Figure 1), of which 21 are permanent grids for long-term monitoring purposes. These represent a total of 1 408 cameras deployed for grids on permanent locations and 873 installed in the roaming locations. More than 43 000 volunteers have annotated approximately 18 months of data for 19 of these sites and classified more than 2 million photos to date.

Our grids and collaborative research allow us to combine different questions and leverage the full potential of camera traps. For example, in the Kruger National Park we have set up the cameras along with studies of the phenology of vegetation cover (e.g. tree and herbaceous composition, cover and structure), which has created an opportunity to monitor the vegetation dynamic while accounting for mammal presence. Further, by-catch data (not targeted species), such as birds or human activity, will represent an interesting opportunity to investigate other species in the future. The potential role of this by-catch data of camera trap studies to conservation efforts or ecological studies has recently been highlighted.[23]

Our grids have documented sightings of some threatened and elusive species outside known distribution ranges as well as unusual behavioural interactions. In 2019, for example, we detected a leopard (*Panthera pardus*) near Karoo National Park. Similarly, we photographed a brown hyaena (*Parahyaena brunnea*) in Camdeboo National Park. None of these records was expected as no official or anecdotal records have existed since colonial times and therefore these sightings represent significant recent geographic range shifts for these species. In June 2019 in the Karoo National Park, one of the Snapshot cameras took 21 pictures of three to five meerkats (*Suricata suricatta*) and a yellow mongoose (*Cynictis penicillata*) foraging together. An analysis of the photographic sequence suggested they were sharing vigilance and staying near each other. These observations constitute new evidence to support the only previous observation of this cooperative behaviour in the Andries Vosloo Kudu Nature Reserve in the Eastern Cape Province.[24]

We aim to be as collaborative as possible and share data to facilitate other conservation and research projects. For example, we have shared ~5500 records with the MammalMAP initiative, which provides a platform for citizen scientists and researchers to contribute biodiversity information for South Africa (http://vmus.adu.org.za/). We expect to make data available on other platforms such as the Foundational Biodiversity Information Programme in South Africa (http://fbip.co.za/) and the Global Biodiversity Information Facility to help evaluate progress toward governmental agreements such as the UN Sustainable Development Goals. Evaluating trends at the broad geographic scale of the Snapshot network can inform IUCN Red List entries on species for which we currently lack accurate estimates of extant population sizes.

Systematic long-term monitoring is crucial to understand the trends in populations and species through time and to facilitate informed management decisions. However, data collection via camera traps has increased at a much faster rate than have our technical capabilities to analyse such large data sets.[25] Another limitation is the access to research groups and computational centres with sufficient processing infrastructure to train and run machine learning algorithms. In our project, for example, the information is analysed by the Lion Center at the University of Minnesota. All data are sent to that lab group for the management of the online classification process and data curation. This can create a bottleneck and slow the data pipeline. One potential solution is to set up multiple hubs within a network, which we plan to enact by copying existing infrastructure onto South Africa's supercomputer at the Centre for High Performance Computing. This continues the trend of distributed management and provides additional opportunities for training and professional development.

Future directions and conclusions

The Snapshot collaborators in South Africa meet annually to review current projects, prioritise future research, and plan funding activities. At the 2019 meeting, we defined five research themes:

- **Theme 1:** The role of anthropogenic landscapes in shaping biodiversity, distributions, populations, and communities.
- **Theme 2:** Investigating ecological interactions and food webs through space and time.
- **Theme 3:** Understanding and predicting the consequences of climate change for mammal behaviour, distribution, adaptation, and community composition.
- **Theme 4:** Assessment of conservation priorities and protected area effectiveness.
- **Theme 5:** Merging camera trap data with other data to address more specific questions and improve monitoring.

Snapshot's integration of scientists, citizens, and technology can provide meaningful assessments of the status of southern African mammals at fine or broad scales. Our collaboration also holds the potential to contribute to advancing statistical and technological capacity in South Africa as well as to support environmental education for the general public. With the current advances in data processing, we are hopeful that Snapshot will provide timely recommendations and relevant ecological information to support southern Africa's national and international commitments to conserving one of the most biodiversity rich regions in the world. However, only political will and multi-sectoral commitments will make it possible to leverage the full potential of technology to produce practical effects.

Reserves and researchers who wish to join the Snapshot network should contact Sarah Huebner at huebn090@umn.edu or Jan Venter at Jan.Venter@mandela.ac.za

Acknowledgements

We thank all the Zooniverse volunteers who contribute classifications to Snapshot Safari, and the moderators who donate their time and expertise to our projects. We also thank sponsors of our work, including the South African National Biodiversity Institute (SANBI), Foundational Biodiversity Information Programme (FBIP), South African National Research Foundation (NRF), Fynbos Trust, Nelson Mandela University, Fairfields, Detroit Zoological Society, Zoo Miami, Cincinnati Zoo Angel Fund, Seneca Park Zoo, and The Living Desert. We thank the Minnesota Supercomputing Institute for providing resources for data storage and processing, among others. Finally, we thank all the people and institutions participating in Snapshot South Africa, including the reserve managers and owners for providing access and supporting this programme, National Parks Institutions of every country, students and volunteer groups helping to maintain the grids, and the governmental institutions overseeing these parks and reserves.

References

1. Díaz S, Settele J, Brondízio ES, Ngo HT, Agard J, Arneth A, et al. Pervasive human-driven decline of life on Earth points to the need for transformative change. Science. 2019;366(6471), eaax3100. https://doi.org/10.1126/science.aax3100

2. Ceballos G, Ehrlich PR, Dirzo R. Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. Proc Natl Acad Sci USA. 2017;114(30):E6089–E6096. https://doi.org/10.1073/pnas.1704949114

3. Ahumada JA, Hurtado J, Lizcano D. Monitoring the status and trends of tropical forest terrestrial vertebrate communities from camera trap data: A tool for conservation. PLoS ONE. 2013;8(9), e73707. https://doi.org/10.1371/journal.pone.0073707

4. Donlan CJ, Tershy BR, Campbell K, Cruz F. Research for requiems: The need for more collaborative action in invasive species management and conservation. Conserv Biol. 2003;17(6):1850–1851. https://doi.org/10.1111/j.1523-1739.2003.00012.x

5. Barnard P, Altwegg R, Ebrahim I, Underhill LG. Early warning systems for biodiversity in southern Africa – How much can citizen science mitigate imperfect data? Biol Conserv. 2017;208:183–188. https://doi.org/10.1016/j.biocon.2016.09.011

6. Jensen D, Campbell J. Discussion paper: The case for a digital ecosystem for the environment: Bringing together data, algorithms and insights for sustainable development. Science Policy Business Forum. Nairobi: UN Environment; 2019.

7. Tilman D, Clark M, Williams DR, Kimmel K, Polasky S, Packer C. Future threats to biodiversity and pathways to their prevention. Nature. 2017;546(7656):73–81. https://doi.org/10.1038/nature22900

8. Child M, Roxburgh L, Do Linh San E, Raimondo D, Davies-Moster H. The 2016 red list of mammals of South Africa, Swaziland and Lesotho. South Africa: South African National Biodiversity Institute, Endangered Wildlife Trust; 2016.

9. Swanson A, Kosmala M, Lintott C, Simpson R, Smith A, Packer C. Snapshot Serengeti, high-frequency annotated camera trap images of 40 mammalian species in an African savanna. Sci Data. 2015;2(1):150026. https://doi.org/10.1038/sdata.2015.26

10. Willi M, Pitman RT, Cardoso AW, Locke C, Swanson A, Boyer A, et al. Identifying animal species in camera trap images using deep learning and citizen science. Methods Ecol Evol. 2019;10(1):80–91. https://doi.org/10.1111/2041-210X.13099

11. Swanson A, Kosmala M, Lintott C, Packer C. A generalized approach for producing, quantifying, and validating citizen science data from wildlife images. Conserv Biol. 2016;30(3):520–531. https://doi.org/https://doi.org/10.1111/cobi.12695

12. Norouzzadeh MS, Nguyen A, Kosmala M, Swanson A, Palmer MS, Packer C, et al. Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. Proc Natl Acad Sci USA. 2018;115(25):E5716–E5725. https://doi.org/10.1073/pnas.1719367115

13. Gomez A, Salazar A, Vargas F. Towards automatic wild animal monitoring: Identification of animal species in camera-trap images using very deep convolutional neural networks. Ecol Inform. 2016;41:24–32. https://doi.org/10.1016/j.ecoinf.2017.07.004

14. Spiers H, Swanson A, Fortson L, Simmons BD, Trouille L, Blickhan S, et al. Everyone counts? Design considerations in online citizen science. J Sci Commun. 2019;18(1):A04. https://doi.org/10.22323/2.18010204

15. Anderson TM, White S, Davis B, Erhardt R, Palmer M, Swanson A, et al. The spatial distribution of African savannah herbivores: Species associations and habitat occupancy in a landscape context. Philos Trans R Soc B Biol Sci. 2016;371(1703), 20150314. https://doi.org/10.1098/rstb.2015.0314

16. Hepler SA, Erhardt R, Anderson TM. Identifying drivers of spatial variation in occupancy with limited replication camera trap data. Ecology. 2018;99(10):2152–2158. https://doi.org/10.1002/ecy.2396

17. Palmer MS, Fieberg J, Swanson A, Kosmala M, Packer C. A 'dynamic' landscape of fear: Prey responses to spatiotemporal variations in predation risk across the lunar cycle. Ecol Lett. 2017;20(11):1364–1373. https://doi.org/10.1111/ele.12832

18. Allen ML, Peterson B, Krofel M. No respect for apex carnivores: Distribution and activity patterns of honey badgers in the Serengeti. Mamm Biol. 2018;89:90–94. https://doi.org/10.1016/j.mambio.2018.01.001

19. Palmer MS, Packer C. Giraffe bed and breakfast: Camera traps reveal Tanzanian yellow-billed oxpeckers roosting on their large mammalian hosts. Afr J Ecol. 2018;56(4):882–884. https://doi.org/10.1111/aje.12505

20. Swanson A, Arnold T, Kosmala M, Forester J, Packer C. In the absence of a "landscape of fear": How lions, hyenas, and cheetahs coexist. Ecol Evol. 2016;6(23):8534–8545. https://doi.org/10.1002/ece3.2569

21. Cusack JJ, Swanson A, Coulson T, Packer C, Carbone C, Dickman AJ, et al. Applying a random encounter model to estimate lion density from camera traps in Serengeti National Park, Tanzania. J Wildl Manage. 2015;79(6):1014–1021. https://doi.org/10.1002/jwmg.902

22. Dickman AJ, Hazzah L, Carbone C, Durant SM. Carnivores, culture and "contagious conflict": Multiple factors influence perceived problems with carnivores in Tanzania's Ruaha landscape. Biol Conserv. 2014;178:19–27. https://doi.org/10.1016/j.biocon.2014.07.011

23. Hofmeester TR, Young S, Juthberg S, Singh NJ, Widemo F, Andrén H, et al. Using by-catch data from wildlife surveys to quantify climatic parameters and the timing of phenology for plants and animals using camera traps. Remote Sens Ecol Conserv. 2019;6(2):129-140. https://doi.org/10.1002/rse2.136

24. Do Linh S, Somers MJ. Mongooses on the move : An apparent case of interspecific cooperative vigilance between carnivores. South Afr J Wildl Res. 2006;36(2):201–203.

25. Ahumada JA, Fegraus E, Birch T, Flores N, Kays R, O'Brien TG, et al. Wildlife insights: A platform to maximize the potential of camera trap and other passive sensor wildlife data for the planet. Environ Conserv. 2019;47(1):1–6. https://doi.org/10.1017/s0376892919000298

Engagement for airborne geophysical survey within a transdisciplinary baseline programme in the Eastern Cape Karoo

AUTHORS:
Barry Morkel[1,2]
Moctar Doucouré[1,3]
Martin Bentley[1,3]
Divan Stroebel[1,3]
Maarten de Wit[1]

AFFILIATIONS:
[1]Africa Earth Observatory Network (AEON)-Earth Stewardship Science Research Institute (ESSRI), Nelson Mandela University, Port Elizabeth, South Africa
[2]Department of History and Political Studies, Nelson Mandela University, Port Elizabeth, South Africa
[3]Department of Geosciences, Nelson Mandela University, Port Elizabeth, South Africa

CORRESPONDENCE TO:
Barry Morkel

EMAIL:
barry.morkel@mandela.ac.za

HOW TO CITE:
Morkel B, Doucouré M, Bentley M, Stroebel D, De Wit M. Engagement for airborne geophysical survey within a transdisciplinary baseline programme in the Eastern Cape Karoo. S Afr J Sci. 2021;117(1/2), Art. #8515. https://doi.org/10.17159/sajs.2021/8515

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
engaged scholarship, airborne geophysical survey, Karoo shale gas, baseline research, social licence

PUBLISHED:
29 January 2021

The biggest lesson from the recent shale gas boom in the USA – specifically in relation to monitoring and mitigating contamination of water, and community health impacts reported in receiving communities, such as the Marcellus shale regions of Pennsylvania – is the fundamental need for baseline studies prior to any proposed shale gas development.[1(p.547)] The USA is the largest (shale) gas producer and exporter to the European Union (EU) but it is generally accepted that, in most of the basins and shale plays in the USA in which unconventional oil and gas development has occurred, baseline studies have not been implemented.[2(p.40)] This phenomenon, along with its environmental monitoring and regulatory implications, led to the establishment of the Karoo Shale Gas Baseline Programme, undertaken by the Africa Earth Observation Network (AEON) – Earth Stewardship Science Research Institute (ESSRI), hereon refered to as AEON, in partnership with the Eastern Cape Provincial Government. The Programme was the first of its kind in South Africa, and produced a transdisciplinary baseline of the Karoo prior to any shale gas development in the Basin.

A report by AEON summarises 4 years of baseline data collection and analyses, and shows that all aspects of life and natural systems in the Karoo could be affected by shale gas development. To provide this data, AEON research teams from Nelson Mandela University conducted groundwater hydrocensus and sampling for chemical analysis, surface water studies and ecosystems analysis, micro-seismicity measurements, and methane gas analysis, socio-economic research, and citizen science.[3]

As the great shale gas debate[4] affects all inhabitants of the Karoo, including farmers, landowners and residents in Karoo towns, permission to conduct natural baseline research calls for meaningful consultation involving deliberate engagement with affected communities based on information sharing, and responsiveness to the specific issues raised by local stakeholders in the process. This process requires greater transparency and trust building than has been evident in the overall Karoo shale gas development process to date.[3-6] Difficulties include the application and utilisation of new and/or relatively unfamiliar data gathering methods and technologies, including fixed-wing airborne and unmanned aerial vehicle (UAV or drone) observations within an environment already experiencing considerable uncertainty and trust deficit amongst many stakeholders.

The focus of this Commentary is stakeholder consultation as well as operational and logistical resilience, both in the field and behind the scenes, that have been necessary to conduct an airborne survey in the Karoo.

Landowner suspicions and trust building

Much of our ground and geophysical baseline data sets cover an area near the Eastern Cape Karoo town of Jansenville, a region of the semi-arid Karoo colloquially referred to as the Noorsveld, located 190 km to the north of Port Elizabeth. This area was prioritised mainly due to groundwater and logistical considerations, as well as previous scientific research undertaken by the group, which included studies of black shales from drillholes, deep conductive layer and velocity anomalies.[7-9] This led to the deployment of different geophysical methods aimed at better quantifying the deep subsurface, such as magneto-telluric surveys conducted in partnership with the GFZ-Potsdam[3], and the application of a novel passive seismic technique (i.e. no induced energy) which was only available through AEON-ESSRI at the time. Airborne data collection was necessary in order to bridge the gap between surface and deep subsurface geophysics.

AEON had developed considerable experience in the area through related baseline studies and field work conducted since mid-2014. These studies were dependent on a productive working relationship with stakeholders, including the local municipality, relevant provincial government departments (through their regional offices located either in Jansenville or Graaff-Reinet) as well as the local business and farming community.

Initial consultations in the survey area primarily focused on providing the local stakeholders of the then Ikwezi Local Municipality (an area now located within the larger Dr Beyers Naude Local Municipality) with an introduction to the Karoo Shale Gas Baseline Research Programme, in order to facilitate initial access to various study sites around the Karoo (see timeline in Figure 1). In Jansenville and across the Noorsveld, prior to commencing with airborne geophysics, the group had already conducted numerous baseline studies, with some still ongoing at the time, including a groundwater survey of 52 boreholes[3,10], a magneto-telluric ground survey[3,8], surface ecosystems monitoring in water bodies and termite mounds[3,11-13], micro-earthquake surveys[5,9], as well as a household survey[13] in and around Jansenville. On implementation of the magneto-telluric survey, a year before commencing preparations for the airborne geophysics, the team convened separate meetings with local farmers' associations, the Ikwezi Municipal Council and senior officials, to inform them in greater detail of the need for the magneto-telluric study and to officially welcome visiting German science collaborators (magneto-telluric survey team) from the GFZ-Potsdam. The team relied heavily on the knowledge and experience of the local farmers' associations (Jansenville, Klipplaat and Waterford) in identifying the many landowners and obtaining their contact details, so as to facilitate direct contact with the individuals involved. Two retired farmworkers, who are familiar with the terrain, assisted the ground team in navigating between the various farms. These contacts proved essential during the roll-out of the magneto-telluric survey in late 2014, because it required extensive interaction with farmers and landowners in gaining access to farms within the study area, and laid a strong foundation for the airborne survey.

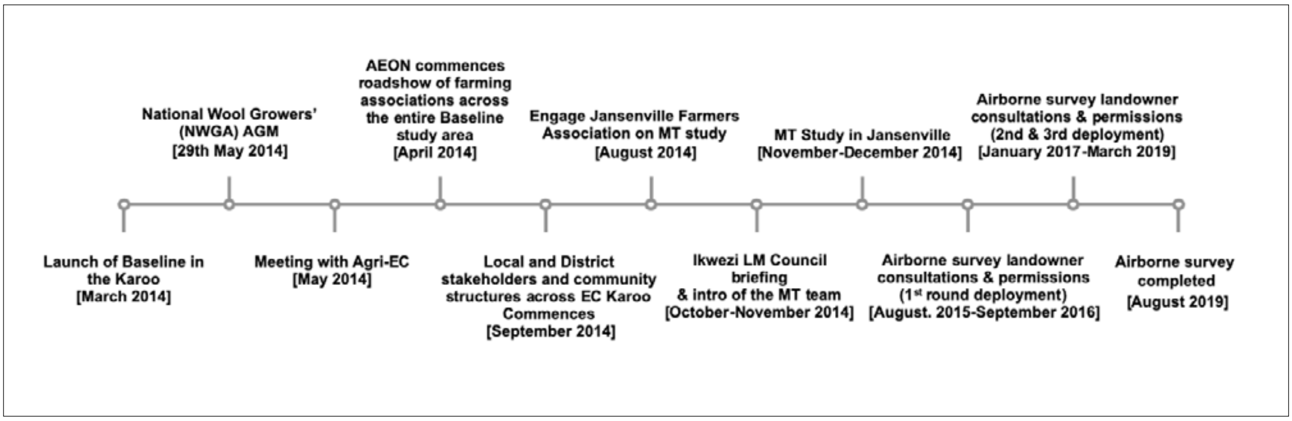


Figure 1: Stakeholder engagement and fieldwork timeline for the Karoo Baseline and Jansenville Airborne Surveys.

Using the same strategies adopted for the smaller magneto-telluric survey, the ground team conducted a farm-to-farm campaign as best as it could. In instances in which the farmer or landowner did not attend our meetings, which were convened with farm associations in the District, the team visited the farm in person or phoned the owner to let them know of the survey and to obtain permission to fly over their property. Owing to the complex nature of landownership, land tenure, land use patterns and agricultural activities in the district, the team encountered a number of unforeseen challenges, and learned lessons along the way. In addition to these difficulties related to contacting landowners, areas to be avoided had to be identified. These ultimately resulted in exclusions and avoidances on the survey flight plan so as to avoid negative impacts on the ground. We had to avoid pre-identified areas related to livestock farming (mainly linked to lambing season), wildlife and game breeding (including for hunting and eco-tourism), as well as chicken and ostrich farms because these were sensitive for a number of farmers and landowners. While these presented difficulties for the permission process, these challenges were overcome by the team undertaking 'ground-truthing' missions and plotting coordinates at specific camps and sites in collaboration with landowners who had identified areas of sensitivity, which were excluded from the survey flight plan. This collaboration between ground teams and landowners ultimately resulted in the survey being successfully flown over 141 land parcels (refer to Box 1).

Operationalising the airborne survey in the Karoo

Survey flying commenced in October 2016 with the use of a gyrocopter. The team decided that a gyrocopter was the platform of choice because it is less invasive, with little disturbance on the ground, whilst maintaining fuel efficiency. The initial plan was to collect data by flying at an altitude of approximately 40 m, across a grid of N-S flight lines spaced 200 m apart at a speed of 80 km/h. Shortly after commencement, and on review of the initial data collected, flights were suspended owing to strong winds together with the effects of thermal lift during the Karoo's summer months, which had impacted data quality negatively. The survey recommenced in April 2017; this was also cut short after the airborne team experienced a technical malfunction with the aircraft which was left irreparable after an emergency landing.

Given the difficulty in sourcing a replacement gyrocopter, it was then decided to use a fixed-wing Maule Bush aircraft which offered greater stability in the air. However, this new aircraft necessitated further consultation and communication on the ground. An advertisement was published in the local newspaper, the *Graaff-Reinet Advertiser,* and posters were displayed at high visibility sites in Jansenville (library, police stations, shops, municipal offices and local farmers' cooperative). The decision was also taken to re-schedule flying for cooler winter months, thereby ensuring more favourable wind conditions for better data quality.

The first deployment of the fixed-wing aircraft was cut short in 2018 by damage to an essential sensor. The team then finally began the survey again in June 2019. To avoid those areas identified as 'Not allowed' and 'Owner not contactable' (Figure 2), the airborne team had to implement necessary technical measures during planning and the flight itself. Similarly, over farms where it was known that hunting occurred (see Box 1), the height above the ground was increased to 80 m because flying continued during the colder hunting season in the Karoo. While the increase in flight height above ground had the effect of lowering the spatial definition of detectable magnetic features, the gaps in the survey coverage mean that some subtle local structures were unavailable to the interpreter of the geophysical image.

Box 1: Conditions on the ground that make for a dynamic and fluid environment for research teams in the Karoo

The team also liaised with the regional Conservation Management Office of the Provincial Department of Economic Development and Environmental Affairs, located in Graaff-Reinet, in order to better understand the extent and complexities associated with the game breeding and hunting industry in the district due to their economic importance. Compounding the situation was the impact of the prolonged drought experienced in recent years, with a number of farmers raising concerns about flying over certain areas during the harsh summer of the lambing season.

Whilst consultations and preparatory work proved time-consuming, it made a considerable difference in the field, leading initially to the identification of 206 individual land parcels, owned by 115 unique landowners. Based on these initial engagements, our team received permission to fly over 164 of these land parcels; we were denied permission to fly over 15 land parcels, and further excluded another 30 due to the landowner being uncontactable or unidentified by the ground teams (see Figure 2). Subsequent to initial deployments in 2016/2017, and after further adjustments to the survey flight plan, the total final area surveyed included 141 land parcels across approximately 1606 km^2. It was also important that we maintained communication with the then Ikwezi Local Municipality (now Dr Beyers Naude Local Municipality) during preparations for the airborne survey as we required permission to use the local airstrip. Beyond formal permission by Council, ongoing cooperation and support of the relevant officials in the municipality's technical service division remained critical.

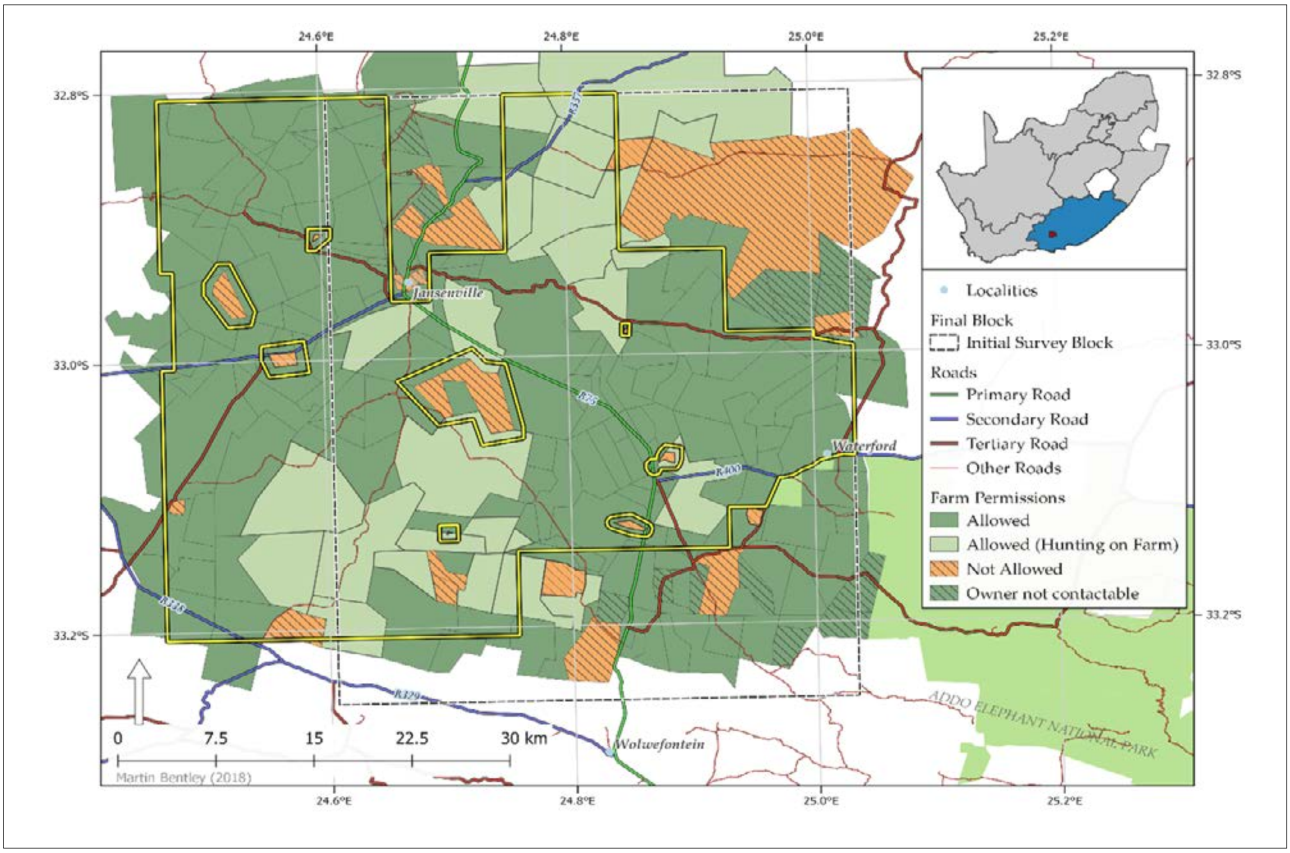


Figure 2: Airborne survey area, showing the coverage obtained during the community consultations and engagements.

Conclusion

The survey, which was eventually concluded in August 2019, has contributed valuable geophysical data to the Karoo baseline study. In addition, the team learnt about negotiating with landowners, farmers and other stakeholders whilst preparing to undertake this airborne survey. In particular, we were aware of the context of distrust that is connected with any research relating to shale gas in the Karoo.

We have shown that permitting for geophysical airborne surveys in the Karoo requires trust building through engagement, and responsiveness to the sensitivities and issues raised by stakeholders in the area. Participatory science, in the form of citizen science, can also prove a significant contribution to trust building. More fundamentally, the operationalisation of airborne surveys in the Karoo as part of the broader natural baseline study has shown that any future exploration for shale gas in the Karoo would require a social licence to operate, which will be reliant on significantly higher levels of consultation and trust building. In the Karoo, where the latter is in short supply, the prevailing trust deficit will only be overcome through honest, open and responsive communication with interested and affected parties moving forward.

During the compilation of this article, one of the authors, AEON Science Director and mentor, Prof. Maarten de Wit, passed away. His passion for science and innovation in fieldwork and data collection through the application of new technologies in the Karoo has been groundbreaking. Similarly, his respect and genuine concern for the people of the Karoo was translated into their inclusion and participation becoming central to the baseline study.

Acknowledgements

This engagement work was made possible by funding from the Eastern Cape Provincial Government, and from the NRF/DSI through Iphakade. The airborne survey was conducted by GyroLAG. This is AEON publication number 198.

Competing interests

We declare that there are no competing interests.

References

1. King R. We need a fracking baseline. La L Rev. 2016;77(2), Art. #12. Available from: https://digitalcommons.law.lsu.edu/lalrev/vol77/iss2/12

2. US Environmental Protection Agency. Hydraulic fracturing for oil and gas: Impacts from the hydraulic fracturing water cycle on drinking water resources in the United States. Executive summary. EPA/600/R-16/236ES. Washington DC: Office of Research and Development; 2016. Available from: https://www.epa.gov/hfstudy/executive-summary-hydraulic-fracturing-study-final-assessment-2016

3. AEON-ESSRI Research Group. (T)(C)racking the shale-gas debate: Baseline data and citizen science. Port Elizabeth: Africa Earth Observatory Network (AEON), Nelson Mandela University; 2018. Available from: https://aeon.org.za/wp-content/uploads/2018/12/AEON-Shale-Gas-Report-2018.pdf

4. De Wit MJ. The great shale debate in the Karoo. S Afr J Sci. 2011;107(7–8), Art. #791. https://doi.org/10.4102/sajs.v107i7/8.791

5. Academy of Science of South Africa (ASSAf). South Africa's technical readiness to support the shale gas industry. Pretoria: ASSAf; 2016. http://dx.doi.org/10.17159/assaf.2016/0003

6. Morkel B, De Wit MJ. Trying not to repeat history: Are shale gas development and broad-based economic development compatible? In: Gamper-Rabindran S, editor. The shale dilemma: A global perspective on fracking and shale development. Pittsburgh, PA: University of Pittsburgh Press; 2018. p. 342–378.

7. Geel C, Schulz HM, Booth P, De Wit M, Horsfield B. Palaeo-environment, diagenesis and characteristics of Permian black shales in the Lower Karoo Supergroup flanking the Cape Fold Belt near Jansenville, Eastern Cape, South Africa: Implications for the shale gas potential of the Karoo Basin. S Afr J Geol. 2015;118(3):249–274. https://doi.org/10.2113/gssajg.118.3.249

8. Weckmann U, Jung A, Branch T, Ritter O. Comparison of electrical conductivity structures and 2D magnetic modelling along two profiles crossing the Beattie Magnetic Anomaly, South Africa. S Afr J Geol. 2007;110(2–3):449–464. https://doi.org/10.2113/gssajg.110.2-3.449

9. Bezuidenhout L, Mordret A, Doucouré M, Lecocq T, Brenguier F, Wagner V, et al. Rayleigh group velocity extraction from ambient seismic noise to map the south Eastern Cape Karoo region. S Afr J Geol. 2017;120:341–350. https://doi.org/10.25131/gssajg.120.3.341

10. Stroebel DH, Thiart C, De Wit MJ. Towards defining a baseline status of scarce groundwater resources in anticipation of hydraulic fracturing in the Eastern Cape Karoo, South Africa: Salinity, aquifer yields and groundwater levels. In: Ofterdinger U, Macdonald AM, Comte JC, Young ME, editors. Groundwater in fractured bedrock environments. Geological Society, London, Special Publications. 2018;479:129–145. https://doi.org/10.1144/SP479.3

11. Mabidi A, Bird MS, Perissinotto R. Assessment of the physicochemical characteristics of surface waterbodies in a region earmarked for shale gas exploration (Eastern Cape Karoo, South Africa). Mar Freshwater Res. 2017;68(9):1626–1641. https://doi.org/10.1071/MF16102

12. Mabidi A, Bird MS, Perissinotto R. Distribution and diversity of aquatic macroinvertebrate assemblages in a semi-arid region earmarked for shale gas exploration (Eastern Cape Karoo, South Africa). PLoS ONE. 2017;12(6), e0178559. https://doi.org/10.1371/journal.pone.0178559

13. Linol B, Miller W, Rensburg C, Schoeman R, Bezuidenhout L, Genin F, et al. Earth stewardship science – Transdisciplinary contributions to quantifying natural and cultural heritage of southernmost Africa. Remote Sens. 2020;12(3):420. https://doi.org/10.3390/rs12030420

AUTHORS:
M. Timm Hoffman[1]
Richard M. Cowling[2,3]
Hana Petersen[1]
Cherryl Walker[4]

AFFILIATIONS:
[1]Plant Conservation Unit, Department of Biological Sciences, University of Cape Town, Cape Town, South Africa
[2]Department of Botany, Nelson Mandela University, Port Elizabeth, South Africa
[3]African Centre for Coastal Palaeoscience, Nelson Mandela University, Port Elizabeth, South Africa
[4]DSI/NRF SARChI Chair in the Sociology of Land, Environment and Sustainable Development, Department of Sociology and Social Anthropology, Stellenbosch University, Stellenbosch, South Africa

CORRESPONDENCE TO:
Michael Hoffman

EMAIL:
timm.hoffman@uct.ac.za

HOW TO CITE:
Hoffman MT, Cowling RM, Petersen H, Walker C. Karoo research update: Progress, gaps and threats. S Afr J Sci. 2021;117(1/2), Art. #8695. https://doi.org/10.17159/sajs.2021/8695

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
bibliography, drylands, interdisciplinarity, publication trends, research gaps

PUBLISHED:
29 January 2021

Karoo research update: Progress, gaps and threats

It has been more than three decades since the conclusion of the Karoo Biome Project (KBP).[1] At its height in the late 1980s, the KBP coordinated the efforts of nearly 100 research projects across a range of mainly ecological and agricultural disciplines. In this brief update we examine the research that has occurred in the Nama-Karoo and Succulent Karoo biomes since then and describe the relative contributions made by different disciplines to this body of knowledge. We also highlight efforts to synthesise knowledge across the disciplinary divides. Finally, we identify notable gaps in the research, especially considering the major land-use changes that are occurring across the Karoo. We conclude that new questions should be asked and that significantly greater collaboration between disciplines should be fostered in order to address the pressing challenges facing the Karoo more effectively. This necessitates a far more coordinated response than has been the case to date. Institutional leadership and additional funding will also be required to achieve this.

Growth and disciplinary focus in the published Karoo literature

To identify the research that has taken place in the Karoo, we searched the Web of Science for all articles using the words Karoo, Karroo, Namaqualand, Richtersveld, Sperrgebiet, Bushmanland, Knersvlakte or Augrabies in their titles, keywords or abstracts. The 5277 articles identified from this search were then reviewed separately by two of the authors (M.T.H. and H.P.). Articles which extended beyond the Karoo region, narrowly defined as the Nama-Karoo and Succulent Karoo biomes[2],were not considered further. Articles for which an abstract was not available were also excluded. The remaining 1578 journal articles (~30% of the original list) were then each assigned a keyword to reflect the primary disciplinary focus.

The selection criteria for our bibliography meant that several important books, book chapters, articles in non-peer reviewed journals, field records and short research notes that are either not indexed in the Web of Science or do not meet our full selection criteria fell out of the analysis. While this is a limitation, particularly with respect to the human sciences, we nevertheless consider that this database provides a broadly indicative and useful overview of the state of Karoo studies, one which can be expanded through follow-up work.

Results show that there has been a steady increase over time in the number of publications concerned with the Karoo (Figure 1a). The last decade of the 20th century was a clear turning point for Karoo research. More than four times the number of articles were published in the decade 1990–1999 than had been produced in all the years since 1946. The momentum created by the KBP undoubtedly contributed to this surge in publications. The number of publications has increased by 30% or more in each subsequent decade. This suggests an ongoing and vibrant research interest in the Karoo which shows little sign of abating.

Research output is, however, not evenly distributed across disciplines (Figure 1b). For example, the geological and palaeosciences together comprise 19% of all articles in our database while the human sciences (primarily anthropology, sociology and archaeology) make up just 9.5%. Most research (~70%) forms part of a broad environmental focus which includes articles in the biological, agricultural and geographical sciences. Evidently, the study of Karoo environments, their biology, their dynamics and how they are used and have changed over time is where the largest research effort has been expended.

Given our selection criteria, the disciplinary emphasis should be interpreted with some caution. Many non-environmental disciplines are not fully represented in the database. This is not only because of the database (Web of Science) and list of keywords used in the initial search, but also because of the additional criteria for inclusion that were applied to the initial selection. Several journals in the human sciences, for example, do not require abstracts with their articles and were excluded. So too were several articles in the geological and palaeoecological sciences which consider deposits and features over regions far larger than our more narrowly defined Karoo study area. Despite these shortcomings, the list of journal articles examined here is revealing of broad trends, both in terms of the increase in overall output and the relative distribution of disciplines.

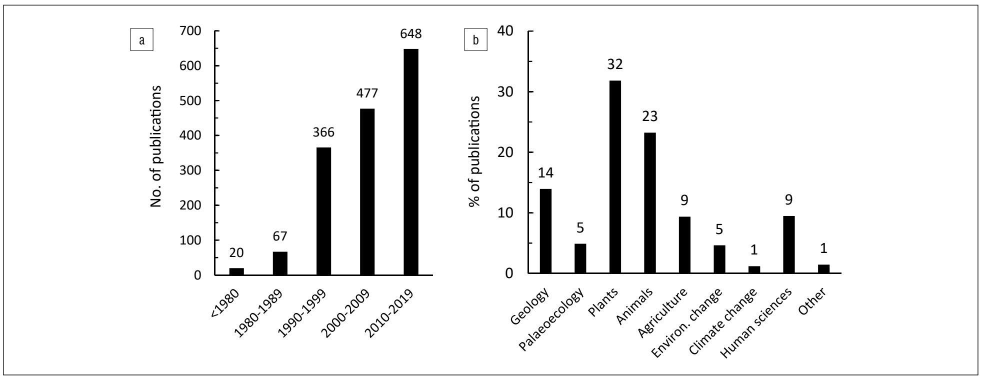


Figure 1: (a) The number of articles (N=1578) concerned with Karoo research and which are listed in the Web of Science for the period 1946–2019. (b) The percentage of publications on the Karoo according to their main disciplinary focus.

Integration: Noteworthy syntheses of Karoo research

The bibliographic analysis highlights the progress that has been made in Karoo studies in specific research disciplines. What effort has been made to synthesise these findings and to integrate knowledge across disciplinary divides? To answer this question, we drew on our own scholarly engagement with Karoo studies spanning several decades.

Two edited book collections (neither of which was captured in our database) and three special issues of peer-reviewed journals have attempted this since 1999, albeit at somewhat different levels of disciplinary integration. The first comprehensive book on Karoo ecology[3] built on the research that had emanated from the KBP. The focus was primarily on a synthesis of what was known about the natural environment at the time. It comprised 20 chapters concerned with the physical environment, the biogeography of the biota as well as the form and function of key plant and animal groups. Chapters on ecological dynamics and the impact of people on the environment were also included. It remains the most important synthesis of the ecology of the entire region to date and several chapters have been cited over 100 times.

The 15 papers in the special issue of *Plant Ecology*, also published in 1999, took Karoo ecological research in a relatively new direction, into the Namaqualand, Richtersveld and Knersvlakte areas of the Greater Cape Floristic Region for the first time.[4] The emphasis was on the diversity, biogeography, physiology and conservation of the flora of the Succulent Karoo biome in relation to key environmental gradients. The impact of grazing and long-term changes in vegetation in response to climate and drought were also included. This was the largest single collection of peer-reviewed ecological research to cover this internationally recognised biodiversity hotspot. It laid the foundation for the subsequent explosion of interest in the region's conservation.

These two syntheses were followed in 2007 by another collection of articles dedicated to the winter rainfall Namaqualand region, published in a special issue of the *Journal of Arid Environments*.[5] Its focus, however, was less on the extraordinary biodiversity of the region and more on pressing management and social issues such as land reform and the contribution of agriculture, remittances and state grants to household livelihoods. The 20 papers in this special issue reflected a relatively new multidisciplinary focus for Karoo studies, with the history, ecology, and sociology of the communal areas in Namaqualand addressed in a single volume for the first time.

One of the longest-running research programmes in the Karoo is BIOTA (Biodiversity Monitoring Transect Analysis) Southern Africa. This initiative was supported by the German Federal Ministry of Education and Research (BMBF) over the period 2000–2010. Its primary focus was on the assessment and monitoring of biodiversity at 37 observatories along a 2000-km transect in South Africa and Namibia. A synthesis of the many outputs of this project by German and southern African scientists is contained in a three-volume set of books, which is freely available online.[6] Measurements at some of the observatories have extended beyond the lifespan of the project to provide valuable insights into long-term changes in plant diversity, especially in response to unusual events such as major droughts.[7]

The need for greater integration of the natural and social sciences has become a regular call in Karoo studies. The most recent synthesis of Karoo research[8] provides the clearest effort to date to understand the region's complex social-ecological systems more holistically. The 22 papers in the special issue of the *African Journal of Range and Forage Science* cover both the Nama-Karoo and Succulent Karoo biomes. Although the natural sciences still dominate, there is a notable presence of the human sciences, with nearly a quarter of the articles drawn from history, archaeology, sociology and anthropology.

Gaps and threats

One important research gap identified through our bibliographic analysis concerns the impact of climate change. Only 20 articles in our database are on this theme, nearly all of which address either changes in climate directly (50%) or the potential impact of future climate change on vegetation (40%). The potential impact of climate change on animals and agriculture is very poorly represented, with only one article listed for each, while broader social impacts are not covered at all. These are critical gaps given the cross-cutting impacts that have been projected for the arid parts of southern Africa as a result of climate change.[9]

Many researchers have also yet to fully appreciate the magnitude of the land-use changes in the Karoo over the last few decades and the need to adjust their research foci accordingly. For example, while commercial agriculture still dominates the landscape, livestock production has declined significantly since the early 1980s. Farm sizes have also increased, and wildlife farming has become more prominent. Relatively little is known about the full extent of such changes and their intersecting social and ecological impacts. The Karoo has also become a major location for the installation of wind and solar energy developments, with some 4% of the combined area of both biomes designated for renewable energy installations.[10] Concerns have been raised about the potentially harmful consequences for biodiversity[11]; their significance for South Africa's energy mix and local social impacts are just beginning to be studied.

Another set of pressures on the Karoo concerns the mining industry's interest in heavy metals and uranium extraction, as well as the targeting of the Nama-Karoo by the fracking industry as a potential source of shale gas. The two main syntheses which address concerns over fracking[12,13] highlight the paucity of information about the likely impacts of this industry on Karoo hydrology and environments. The effect of habitat fragmentation and noise, light and dust pollution created by the preparation and establishment of fracking sites is likely to be extremely consequential for the biota of the Karoo, while the local jobs created are expected to be largely unskilled and short term. Unfortunately, little of the published literature on the Karoo is helpful when trying to predict the impact of such large-scale disturbances on the environment; these developments present unique pressures which demand new studies.

The wide-open spaces and relatively unpolluted skies of the Karoo have also caught the attention of astronomers. The Southern African Large Telescope opened outside Sutherland in 2005 while the world's largest radio telescope, the Square Kilometre Array (SKA), is being constructed near Carnarvon. The environmental impacts are likely to be broadly positive, with the establishment of a national park around the SKA adding considerably to the area under conservation protection in the Nama-Karoo. However, assessing the impact on local social and economic dynamics of the regulatory controls associated especially with radio astronomy is a more complex undertaking.[14]

Final thoughts

There has been an increasing flow of research outputs for the Karoo since 1986. While the interest has been primarily within the environmental sciences, a greater emphasis on the human sciences and interdisciplinary studies is becoming evident. However, much research underestimates the extent to which and significance of how land-use changes have reconstituted the Karoo's social and ecological environments. In this context, knowledge about rangeland ecology and the impact of domestic livestock has relatively limited reach. The new research questions that are emerging also underscore the need for more inter- and cross-disciplinary collaboration.

Even though the Karoo appears peripheral to the major centres of power, it is an historically and ecologically important region that features increasingly prominently in national development plans. Greater investment in Karoo research is urgently needed to advance our understanding and inform policy debates. To be effective, such research needs better coordination and stronger support by stakeholders across the disciplines.

Acknowledgements

This work draws in part on the research supported by the South African Research Chairs Initiative of the Department of Science and Innovation and National Research Foundation of South Africa (grant no. 98765).

Any opinion, finding and conclusion or recommendation expressed in this material is that of the authors and the NRF does not accept any liability in this regard.

Competing interests

We declare that there are no competing interests.

References

1. Cowling RM. Whither Karoo research? S Afr J Sci. 1986;82:409–411.
2. Mucina L, Rutherford MC, editors. The vegetation of South Africa, Lesotho and Swaziland. Strelitzia 19. Pretoria: South African National Biodiversity Institute; 2006.
3. Dean WRJ, Milton SJ, editors. The Karoo: Ecological patterns and processes. Cambridge: Cambridge University Press; 1999.
4. Cowling RM, Esler KJ, Rundel PW. Namaqualand, South Africa – an overview of a unique winter-rainfall desert ecosystem. Plant Ecol. 1999;142:3–21. https://doi.org/10.1023/A:1009831308074
5. Hoffman MT, Allsopp N, Rohde RF. Sustainable land use in Namaqualand, South Africa: Key issues in an interdisciplinary debate. J Arid Environ. 2007;70:561–569. https://doi.org/10.1016/j.jaridenv.2006.11.021
6. Jürgens N, Schmiedel U, Hoffman MT. Structure and elements of the project and course of the project over the three phases. In: Jürgens N, Haarmeyer DH, Luther-Mosebach J, Dengler J, Finckh M, Schmiedel U, editors. Biodiversity in southern Africa. Volume 1: Patterns at local scale – the BIOTA Observatories. Göttingen & Windhoek: Klaus Hess; 2010. p. 2–3.
7. Schmiedel U, Oldeland J. Vegetation responses to seasonal weather conditions and decreasing grazing pressure in the arid Succulent Karoo of South Africa. Afr J Range Forage Sci. 2018;35:303–310. https://doi.org/10.2989/10220119.2018.1531926
8. Henschel JR, Hoffman MT, Walker C. Introduction to the Karoo Special Issue: Trajectories of change in the Anthropocene. Afr J Range Forage Sci. 2018;35:151–156. https://doi.org/10.2989/10220119.2018.1535214
9. Engelbrecht F, Adegoke J, Bopape M-J, Naidoo M, Garland R, Thatcher M, et al. Projections of rapidly rising surface temperatures over Africa under low mitigation. Environ Res Lett. 2015;10, Art. #085004. https://doi.org/10.1088/1748-9326/10/8/085004
10. Hoffman MT, Skowno A, Bell W, Mashele S. Long-term changes in land use, land cover and vegetation in the Karoo drylands of South Africa: Implications for degradation monitoring. Afr J Range Forage Sci. 2018;35:209–221. https://doi.org/10.2989/10220119.2018.1516237
11. Dean WRJ, Seymour CL, Joseph GS. Linear structures in the Karoo, South Africa, and their impacts on biota. Afr J Range Forage Sci. 2018;35:223–232. https://doi.org/10.2989/10220119.2018.1514530
12. Glazewski J, Esterhuyse S, editors. Hydraulic fracturing in the Karoo: Critical legal and environmental perspectives. Cape Town: Juta; 2016.
13. Scholes R, Lochner P, Schreiner G, Snyman-Van der Walt L, De Jager M, editors. Shale gas development in the central Karoo: A scientific assessment of the opportunities and risks. Report no. CSIR/IU/021MH/EXP/2016/003/A. Pretoria: Council for Scientific and Industrial Research; 2016.
14. Walker C, Chinigò D, Dubow S. Karoo futures: Astronomy in place and space – Introduction. J South Afr Stud. 2019;45:627–639. https://doi.org/10.1080/03057070.2019.1654664

AUTHORS:
Molly V. Czachur[1]
Melvi Todd[2]
Tainã Gonçalves Loureiro[3]
James M. Azam[4]
Siphokazi Nyeleka[5]
Amanda Alblas[6]
Sarah J. Davies[3]

AFFILIATIONS:
[1]Department of Botany and Zoology, Stellenbosch University, Stellenbosch, South Africa
[2]Department of Food Science, Stellenbosch University, Stellenbosch, South Africa
[3]DSI-NRF Centre of Excellence for Invasion Biology (CIB), Stellenbosch University, Stellenbosch, South Africa
[4]DSI-NRF Centre of Excellence for Epidemiological Modelling and Analysis (SACEMA), Stellenbosch University, Stellenbosch, South Africa
[5]Department of Animal Science, Stellenbosch University, Stellenbosch, South Africa
[6]Anthropology Research Unit, Division of Clinical Anatomy, Stellenbosch University, Cape Town, South Africa

CORRESPONDENCE TO:
Molly Czachur

EMAIL:
mollyczachur@gmail.com

HOW TO CITE:
Czachur MV, Todd M, Gonçalves Loureiro T, Azam JM, Nyeleka S, Alblas A, et al. Opening the floor for discussion: A perspective on how scholars perceive attitudes to science in policymaking in South Africa. S Afr J Sci. 2021;117(1/2), Art. #7948. https://doi.org/10.17159/sajs.2021/7948

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
science communication, research communication, policy development, decision-making

PUBLISHED:
29 January 2021

Opening the floor for discussion: A perspective on how scholars perceive attitudes to science in policymaking in South Africa

Policymakers are a vital group with whom scientific research should be communicated, especially when the reason for many research projects is linked to relevance for socio-political and economic management. Science communication has a vital role in transforming research into policy, and a core element of this process is understanding the target group, namely policymakers. Science and policy influence each other deeply, so researchers and policymakers should improve their understanding of each other and of the processes involved in both fields in order to better collaborate. Accordingly, an in-depth understanding of how scholars perceive policymakers is a precondition for scientists to achieve any desired management and policy impacts.

In December 2019, six researchers and one research manager from Stellenbosch University, South Africa, gathered to discuss their understanding of policymakers. The discussion was part of a Science Communication Masterclass hosted by the South African Research Chair in Science Communication and Econnect Communication, Australia. The purpose of the group discussion was to develop a science communication strategy that would enhance the sharing of scientific research outputs with policymakers in South Africa. We explored five questions to help us record our perception of policymakers: (1) Who do we think the policymakers are? (2) How do we think that policymakers perceive research? (3) What concerns do we believe that policymakers have about research? (4) What information do we think policymakers are interested in? (5) What forms of communication do we think policymakers prefer?

This Commentary presents our view on how we think researchers perceive policymakers, as discussed in the Science Communication Masterclass. We hope to initiate a discussion around science communication with policymakers, and improve current practices.

Describing policymakers

The term 'policymaker' is widely used to refer to people who have political influence that directly develops or changes policies, regulations, rules and directives. In our discussions, we initially identified policymakers as people employed in government agencies and based our views on those with whom we have regularly interacted. Non-government organisations (NGOs) were later included as 'policymakers' owing to their impacts on the actions of some communities and their potential influence on government decision-making. We did not include the private sector in our description of policymakers, because we have limited experience of this scenario, but we acknowledge that the private sector is an important component of decision-making and should be explored further. We also recognise that policymaking is a complex process and that no single individual or group of individuals has absolute control over the drafting, editing, consultation, discussion and adoption of policies.

The group acknowledged that policymakers' networks are from different sectors (e.g. private, government, NGOs); thus identifying and connecting with a person or department as the access point might be challenging for researchers, as well as to further report the scientific findings or shape the information to the specific public.[1]

We bring to this Commentary a range of research experience that broadly encompasses environmental conservation, ethics, health, packaging development, food science, fisheries, and invasive species management. We feel that in our research, we would most likely be communicating with officials in the Ministries of Health and Environment, Forestry and Fisheries. The Commentary is written from our perspective, yet there are likely many unique researcher perspectives in South Africa, so a further discussion on this topic is highly encouraged beyond this Commentary.

The policymaker's perception of research

University scholars and policymakers typically have different goals and are embedded within very different bureaucratic systems.[2] Typically, the two groups value different types of information, are subject to differing time frames, and may use distinct 'languages'. However, there are also many scientists working in science councils and other parastatals, such as the Council for Scientific and Industrial Research (CSIR), South African National Biodiversity Institute (SANBI) and the Medical Research Council (MRC), as well as in government departments, who engage more frequently with policymakers, and in different ways from those experienced by scientists embedded in universities. Recognising these differences, we discussed some of the perceptions that policymakers might have about research and identified several problem areas, the most significant of which are outlined below.

Research can be inaccessible

Policymakers may find primary research material inaccessible. Reasons may be the limitations on (1) the physical availability and accessibility of the primary literature and underlying data (policymakers may not have subscriptions to closed access/paywalled journal articles) and (2) the expertise to interpret and use research outputs to inform their policy-related decision-making[3] and (3) the time and resources required to keep up with a burgeoning and continually evolving body of literature. Indeed, policymakers have acknowledged that they often turn to secondary sources of information.[4,5] Thus, limited access to research may maintain the discovery–delivery gap and hinder the use of relevant research for policymaking.[6]

Research is sometimes not fit for purpose

Moreover, research outputs may differ from expected results and hypotheses, or stray from the project as originally proposed. This may happen as research and knowledge constantly change and evolve, but also because of external influences like unforeseen political, social, and environmental events. These changes in research and context may frustrate policymakers and even result in suspicion and mistrust, especially if research outputs were intended to inform specific policies.

An erosion of mutual trust or respect can deeply affect the ability of different groups to work together and may inhibit collaboration.[2,7] The importance of personal contact between researchers and policymakers has been highlighted by Innvaer et al.[8] and Von der Heyden et al.[3] Personal trust and respectful contact are important enabling factors for the transformation of research into policy and may build lasting connections.

Research can take time

Whether commissioned or not, research often takes a long time and there are often notable time lags between project conception and completion. These factors are of concern to policymakers, who require evidence to support urgent decisions.[3] Policymaking is directly allied to the fluctuations of society and therefore must move more quickly than research. It is also, often, urgent. By the time that research findings are ready to be used in policy, the specific issue may no longer be a priority.[9] If the research information is already published or in the public domain before the policymaker requires the output, then more time may be needed even if no new data are required, as publications may need to be interpreted and aligned with the policy decision-making process.

Research can be expensive

Budget limitations were identified as a major concern to policymakers when considering the expense of accessing published research or of commissioning research to address policy questions. When incorporating research into decision-making, policymakers may be concerned about the high cost involved, especially as the outputs (e.g. publications) may not be readily accessible to them or to their constituencies. Researchers, therefore, should supply cost–benefit or cost-effectiveness comparisons to motivate research support. By clearly defining the benefits of research for policy change, researchers can demonstrate that the research is fit-for-purpose, therefore reducing the perceived financial risk to policymakers.

The research message is often complex

Statistical methods underpinning research findings are often based on estimates, *p*-values, confidence intervals, and other metrics unfamiliar to non-specialists, and the key benefits and limitations of the methods may be poorly presented. Communication of uncertainty is another essential part of scientific discourse, as researchers attempt to characterise and quantify uncertainty (e.g. assumptions) in publications and presentations.[10] The complexity in the discourse around research results, as well as inaccurate communication of uncertainty and study assumptions can result in policymakers receiving wrong or even perverse messages about research findings.[6,9]

In order to support decision-making, policymakers may require research findings to be communicated by summaries in less technical formats, for example, policy briefs. This can present a challenge to researchers, who are seldom trained to present their methods or findings in such formats. However, this task is important for researchers to master if they are to communicate effectively with policymakers.[11,12]

Scalability and relevance

The scalability of research results may also concern policymakers. We understand scalability to be the suitability or adaptability of findings to settings other than the specific one studied. In fields such as health and environmental monitoring, the results of site-specific or demographic-specific studies may not apply to a wider domain beyond the original study area. Given that we argue that policymaking is wide-scale decision-making, we believe that policymakers may well have concerns over whether research projects can be generalised to a broader area.

Policymakers are also impacted by the extent to which researchers involve stakeholders who might be affected by the policy changes. At first, we argued that a high level of stakeholder involvement was a high priority for policymakers. However, where links between research and stakeholders are unclear, policymakers might question the relevance of the research for large-scale policy change. Ideally, co-production of knowledge should be undertaken to ensure that the outcomes are relevant and useful for all role players.[13-15]

What information do policymakers want from researchers?

Policymakers usually want research outcomes that are practical and can be applied directly in policy formulation. Evidence-based research and results are used to defend a policy change or an action. From the point of view of a researcher, the interaction between researcher and policymaker may be (1) passive – the policymaker wants to design a project or needs the help of an expert and contacts the researcher for information, or (2) active – the researcher contacts a policymaker who is in a position to use their work.[11] The communication approach and its outcome might be different, depending on the situation that led to the interaction. In passive interactions, we anticipate that policymakers are already well informed about the research, and it is the communication itself that may present challenges.

When a researcher actively approaches a policymaker, there is a need for detailed explanation of the implications of the research and how it might guide or support policymaking. However, the policymaker may question the validity of the research, as well as its applicability to a wider scenario or context. To have the best results with proactive approaches, researchers need to be clear about what they want to achieve, produce strong evidence therefor, and emphasise the validity of the research findings, as well as the applicability to the wider policy landscape. Finally, based on the cost (time and money) and the scale of potential impacts of amending existing policies, we believe that policymakers place substantial value on how well research results align with existing policy and the political context and address shortcomings that may have been identified in current policy. Research is unlikely to influence policy if the required changes are not feasible in the short term.

What form of communication do policymakers prefer?

We also attempted to identify the forms of communication that policymakers might prefer. The amount of information that can be transferred in a communication interaction is limited, and policymakers receive substantial amounts of information every day from different sources.[9] A study from the USA showed that by far the largest part of any information given to a policymaker is not assimilated or even accessed.[16] The key to effective research communication with policymakers, therefore, is to provide policy-specific (relevant) information free from unnecessary embellishments.

In our experience, the most productive interactions with policymakers are based in personal communication between researcher and policymaker, rather than in broadcast methods such as emails, reports or brochures. Face-to-face interactions such as discussion sessions, telephone calls, or feedback workshops are the most effective and preferred. However, formal written communications, such as reports and policy or media briefs, are often effective in facilitating discussion when they follow or are combined with personal interaction.

For researchers, personal communication with policymakers takes time and effort, and may be more feasible when a researcher has a working knowledge of the policymakers in their field. When there is no prior knowledge of the key players, then some form of relationship-building must first take place. Similarly, if policymakers can identify the key researchers who are generating evidence related to their policy, they may find it easier to approach the researcher to establish a communication channel.

Globally, there have been various attempts to bridge the communication gap between researchers and policymakers, such as the 'Science meets Parliament' initiative in Australia.[17,18] At these gatherings, practising scientists and Members of Parliament discuss mutually agreed priorities. Two positive outcomes include: (1) opening channels of communication between scientists and politicians, and (2) building a cohort of scientists who are comfortable engaging with politicians and the policymaking processes.[17]

Whether or not policymakers hold scientific qualifications, they are expected to understand the often-complex policies they draft and approve as well as the supporting evidence. Similarly, researchers typically do not have policy experience and often struggle to engage with the official processes that are involved. Policymakers and researchers/scholars often have different backgrounds and expertise, increasing the challenges in communication between them. Accordingly, we believe that the implementation of a 'Science meets Parliament' or similar initiative in South Africa might involve a broad range of policymakers, including political representatives and senior staff of government departments who meet face-to-face with researchers. We believe that such action might start to close the communication gap between these important groups and begin to form a community of practice for evidence-based policymaking in South Africa.

Figure 1 summarises this discussion and provides a visual interpretation of our perceptions of policymakers. We aim to encourage follow-up discussions with policymakers and researchers alike.

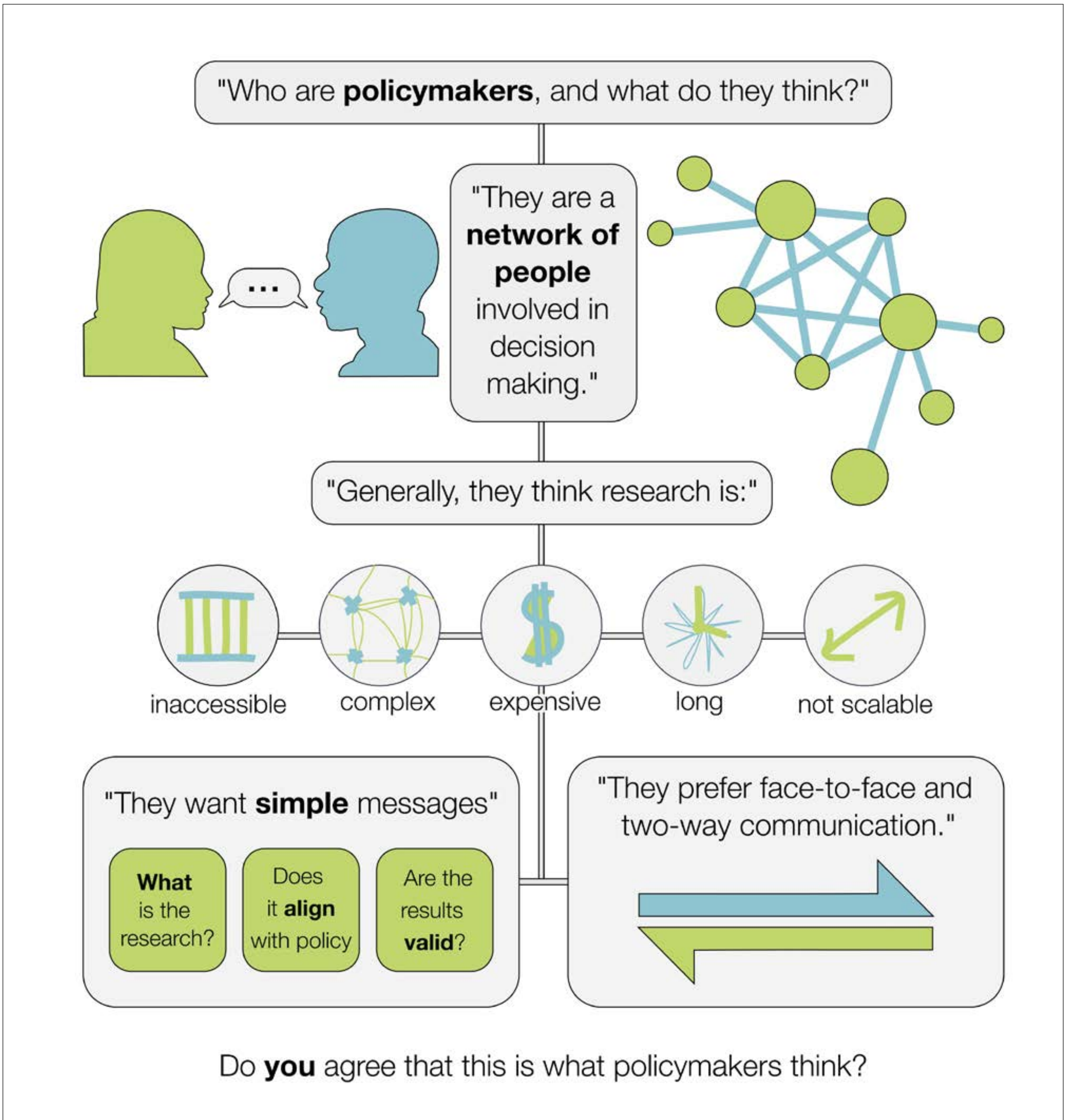


Figure 1: A researcher's perception of policymakers based on discussions held between seven researchers/research managers who met at the Science Communication Masterclass hosted at Stellenbosch University in December 2019.

Conclusions

After defining policymakers and understanding their thinking and their needs, several factors emerged that could hinder effective communication between them and researchers. Only by identifying these hurdles and finding effective measures to address them will we ensure that both groups can effectively achieve their goals.

A major gap we identified in our discussion was that researchers' perception of policymakers may well not represent the opinions and perspectives of policymakers themselves. Science communication is not a one-way transfer of knowledge, but may take many different directions, and policymakers will have their own contributions and experiences to share about researchers. Policymakers should be included in any future discussions to identify how they and researchers can become better connected. It would also be useful to outline exactly what researchers need from policymakers in order to ensure their research outputs are relevant and accessible, to define reciprocally who researchers are, and how researchers could be communicated with more effectively.

Future research needs to untangle the interactions between researcher perceptions of how policies are formulated and changed in practice. Furthermore, policymakers' perspectives should be explored in more detail and be compared to researcher perceptions before any conclusive statements can be made about science communication planning with policymakers. The formation of a more effective dialogue between policymakers and scientists should be addressed urgently to ensure that better understanding is created.

We recommend that an annual, formal interaction between researchers and government officials be facilitated through a 'Science meets Parliament' or similar initiative. Similar events in, Australia[17,18], Canada[19] and the European Union have encouraged better translation of science into policy.

Finally, it is important to bear in mind that a person's perception may be mistaken, biased or context-dependent when considering different fields and research arenas. Ultimately, by expanding this discussion and exploring the interactions between perceptions and policy change, we will be better able to identify major areas where researcher–policymaker communication would be improved. We aim to stimulate debate around the ideas discussed here and encourage solutions to how the research–policy community might work together more effectively.

Acknowledgements

The original discussion outlined here was held during the Science Communication Masterclass on 4 December 2019 at Stellenbosch University, South Africa. The workshop session and guiding questions were led by Dr Marina Joubert, Dr Jenni Metcalfe and Mr Toss Gascoigne. The authors formed a sub-group of the larger Masterclass participants, of which there were 42 participants in total. We thank the organisers of the Masterclass for bringing together researchers to discuss the relevance of target groups, particularly policymakers, in science communication. S.J.D. and T.G.L. are supported by the DSI-NRF Centre of Excellence for Invasion Biology (South African National Research Foundation grant no. 41313). M.V.C. is supported by the South African National Research Foundation. J.M.A. is supported by a DSI-NRF Centre of Excellence for Epidemiological Modelling and Analysis PhD bursary. M.T. is supported by a DAAD-NRF Doctoral scholarship (Grant Reference number: MND190508435061) and a portion of her work forms part of the DSI Honeybush Project (Contract DST/CON 00029/2019).

Competing interests

We declare that there are no competing interests.

References

1. Farquharson K. A different kind of snowball: Identifying key policymakers. Int J Soc Res Meth. 2005;8(4):345–353. https://doi.org/10.1080/1364557042000203116
2. Choi BC, Pang T, Lin V, Puska P, Sherman G, Goddard M, et al. Can scientists and policy makers work together? J Epidemiol Community Health. 2005;59(8):632–637. http://dx.doi.org/10.1136/jech.2004.031765
3. Von der Heyden S, Lukey P, Celliers L, Prochazka K, Lombard AT. Science to policy – Reflections on the South African reality. S Afr J Sci. 2016;112(11/12), Art. #a0183. http://dx.doi.org/10.17159/sajs.2016/a0183
4. Lukey P. Triggering informed action. Paper presented at: The First United Nations World Data Forum; 2017 January 15–18; Cape Town, South Africa.
5. Paine CG, Sadan M. Use of evidence in policy making in South Africa: An exploratory study of attitudes of senior government officials. Afr Eval J. 2015;3(1), Art. #145. http://dx.doi.org/10.4102/aej.v3i1.145
6. Kerner JF Integrating research, practice, and policy: What we see depends on where we stand. J Public Health Manag Pract. 2008;14(2):193–198. https://doi.org/10.1097/01.phh.0000311899.11197.db
7. Isenberg P. The current drought exposes – not creates – long-standing water problems: Can policy-makers and scientists learn from this? San Franc Estuary Watershed Sci. 2014;12(2), Art. #2. https://doi.org/10.15447/sfews.2014v12iss2art2
8. Innvaer S, Vist G, Trommald M, Oxman, A. Health policy-makers' perceptions of their use of evidence: A systematic review. J Health Serv Res Policy. 2002;7:239–244. https://doi.org/10.1258/135581902320432778
9. Brownson RC, Royer C, Ewing R, Mcbride T.D. Researchers and policymakers: Travelers in parallel universes. Am J Prev Med. 2006;30(2):164–172. https://doi.org/10.1016/j.amepre.2005.10.004
10. Fischoff B, Davis AL. Communicating scientific uncertainty. Proc Natl Acad Sci USA. 2014;111(4):13664–13671. https://doi.org/10.1073/pnas.1317504111
11. Choi BCK, Gupta A, Ward B. Good thinking: Six ways to bridge the gap between scientists and policy makers. J Epidemiol Community Health. 2009;63(3):179–180. http://dx.doi.org/10.1136/jech.2008.082636
12. Mwenesi HA. Communities, policy-makers and scientists: A critical partnership for malaria elimination. Malar J. 2012;11(1), Art. #14. https://doi.org/10.1186/1475-2875-11-S1-O14
13. Schlesinger WH. Translational ecology. Science. 2010;329(1):609. https://doi.org/10.1126/science.1195624
14. Norström AV, Cvitanovic C, Löf MF, West S, Wyborn C, Balvanera P, et al. Principles for knowledge co-production in sustainability research. Nat Sustain. 2020;3:182–190. https://doi.org/10.1038/s41893-019-0448-2
15. Enquist CA, Jackson ST, Garfin GM, Davis FW, Gerber LR, Littell JA, et al. Foundations of translational ecology. Front Ecol Environ. 2017;15(1):541–550. https://doi.org/10.1002/fee.1733
16. Sorian R, Baugh T. Power of information: Closing the gap between research and policy when it comes to conveying complex information to busy policy-makers, a picture is truly worth a thousand words. Health Aff. 2002;21(2):264–273. http://dx.doi.org/10.1377/hlthaff.21.2.264
17. Gascoigne T. Science meets parliament in Canberra. Eos. 2001;82(1):515–516. https://doi.org/10.1029/01EO00309
18. Gascoigne TH. Science meets the parliament: Australian case study. Paper presented at: Conference on Communicating European Research; 2005 November 14–15; Brussels, Belgium.
19. Zhao J, Azad M, Bertrand EM, Burton C, Crooks VA, Dawson J, et al. Canadian science meets parliament: Building relationships between scientists and policymakers. Sci Public Policy. 2020;47(2):298. https://doi.org/10.1093/scipol/scz062

Professor D.P. Mason (MASSAF, FRSSAF) on the occasion of his 75th birthday: A lifetime contribution to the development of Applied Mathematics in South Africa

AUTHORS:
Ebrahim Momoniat[1]
Marcia Moremedi[2]
Fazal M. Mahomed[3]

AFFILIATIONS:
[1]Department of Mathematics and Applied Mathematics, University of Johannesburg, Johannesburg, South Africa
[2]Department of Mathematical Sciences, University of South Africa, Pretoria, South Africa
[3]DSI-NRF Centre of Excellence in Mathematical and Statistical Sciences, School of Computer Science and Applied Mathematics, University of the Witwatersrand, Johannesburg, South Africa

CORRESPONDENCE TO:
Ebrahim Momoniat

EMAIL:
emomoniat@uj.ac.za

HOW TO CITE:
Momoniat E, Moremedi M, Mahomed FM. Professor D.P. Mason (MASSAF, FRSSAF) on the occasion of his 75th birthday: A lifetime contribution to the development of Applied Mathematics in South Africa. S Afr J Sci. 2021;117(1/2), Art. #9114. https://doi.org/10.17159/sajs.2021/9114

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
continuum mechanics, Mathematics in Industry Study Group

PUBLISHED:
29 January 2021

David Paul Mason turned 75 in 2019. We mark this occasion by acknowledging the contribution made by him to the development of Applied Mathematics in South Africa.

Mason completed his BSc Honours Degree First Class (Mathematics and Physics) at the University of Glasgow in 1966. He went on to complete a DPhil Degree at Oxford University in 1970 on Hilbert boundary value problems in the kinetic theory of ionised gases (plasmas) under the supervision of David Spence. He joined the University of the Witwatersrand in 1970 as a Lecturer in the Department of Applied Mathematics and Computer Science. He was promoted through the academic ranks up to the position of Ad Hominem Chair of Theoretical Mechanics in July 1986. Mason then took up the Chair of Applied Mathematics in 1991 and retired from that position in 2009. He served as Head of the Department of Applied Mathematics for the period 1988–1990 and of the Department of Computational and Applied Mathematics for the period 1997–2000. He is now an Emeritus Research Professor in the School of Computer Science and Applied Mathematics at the University of the Witwatersrand.

Mason has published over 100 articles in Web of Science journals and has supervised or co-supervised 24 MSc and 12 PhD students to completion. He continues to supervise or co-supervise students.

In 1998, he was elected as a Fellow of the Royal Society of South Africa. Two years later he was awarded a South African Mathematical Society Gold Medal for excellence in research. He received the Vice-Chancellor's Academic Citizenship Individual Award from the University of the Witwatersrand in 2013, and a year later was honoured by the South African Mathematical Society for the Advancement of Mathematics. He was elected a Member of the Academy of Science of South Africa in 2015.

He has organised the annual Mathematics in Industry Study Group meetings since 2004. They are held either at the University of the Witwatersrand or the African Institute for Mathematical Sciences in Cape Town. The Study Group meetings bring together problems from industry and problem-solving skills from universities. Industrial mathematics experts from overseas universities are invited to the Study Group to work collaboratively with South African applied mathematicians and graduate students to develop innovative solutions to the problems posed by industry. Typically, five or six problems are posed each year; since 2004, over 80 new problems of industrial origin have been brought into the mathematical sciences in South Africa. These meetings are unique in that academics – and graduate students – can publish their findings not long after the meeting. The most significant result of these Study Groups has been the number of tangible outputs through research articles, postgraduate students and the publication of conference proceedings.

Mason with participants from the 2009 Mathematics in Industry Study Group.

Mason makes innovative use of applied mathematics as a medium for transformation and human capacity development through his postgraduate student supervision. He also contributes to the development of innovation in problem-solving in the discipline of applied mathematics in universities across South Africa through the Graduate Modelling Camp which is held the week before the Mathematics in Industry Study Group meeting. The Graduate Modelling Camp aims to prepare the graduate students for the Study Group. The graduate students, who are from the universities participating in the Study Group, learn problem-solving skills and skills in scientific communication and leadership by working in small groups on problems from earlier Study Group meetings.

Mason works closely with his higher degree students and jointly authors research articles with them, which has positively impacted on generating well-qualified academic staff. A list of Mason's MSc and PhD graduates who work in academia is given in the table. These past students who are now pursuing academic careers are based at universities across South Africa with a few teaching internationally. Other higher degree students, such as Nicolette Roussos (PhD 2001) and Charlemagne Pooe (MSc 1988) spent time in academia but are now employed in financial institutions and consultancies, where their problem-solving skills are highly valued.

Students supervised by Professor Mason who now work in academia

PhD Students		MSc Students	
E. Momoniat (UJ)	1999	S.D. Maharaj (UKZN)	1984
G.M. Moremedi (Unisa)	2001	J.S. Mathunjwa (Swaziland)	1995
R Naz (Lahore School of Economics, Pakistan)	2008	E. Mubai (Wits)	2019
A.G. Fareo (Wits)	2013		
K.P. Pereira (Wits)	2013		
M.R.R. Kgatle (Wits)	2015		
A.J. Hutchinson (Wits)	2016		
N. Modhien (Wits)	2018		
A.B. Magan (Wits)	2019		

Mason's research area is continuum mechanics. He regards the mechanics of continuous media as a unified discipline with different materials, such as fluids and elastic solids, satisfying the same balance laws for mass, momentum, angular momentum and energy, but with each material having its own constitutive equation. In fluid mechanics, he has contributed to boundary layer theory, thin fluid film theory, slow viscous flow and turbulent shear flows. In solid mechanics, he has applied infinitesimal elasticity to cylindrical excavations in mines, and he has contributed to non-linear oscillations in finite elasticity. He also has contributed to the interaction of fluids with solids, in porous elastic media and in hydraulic fracturing.

He contributed to the formation by Professor Fazal Mahomed of the Centre for Differential Equations, Continuum Mechanics and Applications at the University of the Witwatersrand. This research group was highly regarded both locally and internationally. Through the Centre, the powerful methods of Lie group analysis and conservation laws for differential equations were applied to problems in continuum mechanics. Together with his graduate students, Mason investigated the construction of conservation laws for both Prandtl's boundary layer equations and wake flows behind obstacles.

Mason has made a significant contribution to applied mathematics at the University of the Witwatersrand over five decades. He is one of the main reasons that research in continuum mechanics and applied mathematics has thrived and attracts many young researchers. He is an inspiration for many senior academics. His links to international institutions are key in international researchers visiting South Africa. He has made – and continues to make – a significant contribution to human capacity development in the discipline of Applied Mathematics. He continues to develop and innovate in problem-solving in the discipline of Applied Mathematics.

In honour of Mason's 75th birthday, a special session on fluid dynamics was held at the International Conference on Mathematical Sciences and Applications in 2019.

Mason (right) with Moremedi (middle) and Momoniat (left) at the International Conference on Mathematical Sciences and Applications from 25 to 27 September 2019.

Lightning monitoring and detection techniques: Progress and challenges in South Africa

AUTHORS:
Maqsooda Mahomed[1]
Alistair D. Clulow[1,2]
Sheldon Strydom[2,3]
Michael J. Savage[2]
Tafadzwanashe Mabhaudhi[1,4]

AFFILIATIONS:
[1]Centre for Water Resources Research, School of Agricultural, Earth and Environmental Sciences, University of KwaZulu-Natal, Pietermaritzburg, South Africa
[2]Discipline of Agrometeorology, School of Agricultural, Earth and Environmental Sciences, University of KwaZulu-Natal, Pietermaritzburg, South Africa
[3]Department of Geography and Environmental Sciences, School of Geo and Spatial Sciences, North-West University, Mafikeng, South Africa
[4]Centre for Transformative Agricultural and Food Systems, School of Agricultural, Earth and Environmental Sciences, University of KwaZulu-Natal, Pietermaritzburg, South Africa

CORRESPONDENCE TO:
Alistair Clulow

EMAIL:
Clulowa@ukzn.ac.za

DATES:
Received: 04 Sep. 2019
Revised: 08 July 2020
Accepted: 09 July 2020
Published: 29 Jan. 2021

HOW TO CITE:
Mahomed M, Clulow AD, Strydom S, Savage MJ, Mabhaudhi T. Lightning monitoring and detection techniques: Progress and challenges in South Africa. S Afr J Sci. 2021;117(1/2), Art. #7020. https://doi.org/10.17159/sajs.2021/7020

ARTICLE INCLUDES:
☒ Peer review
☒ Supplementary material

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

EDITOR:
Yali Woyessa

KEYWORDS:
climate, climate change, extreme weather, lightning activity, rural communities

FUNDING:
Durban Research Action Partnership; Department of Environment, Forestry and Fisheries; South African National Biodiversity Institute; uMgungundlovu District Municipality; Centre for Transformative Agricultural and Food Systems, University of KwaZulu-Natal.

Globally, lightning causes significant injury, death, and damage to infrastructure annually. In comparison to the rest of the world, South Africa has one of the highest incidences of lightning-related injuries and deaths. The latest available lightning detection techniques and technologies are reviewed and include current research in South Africa and South Africa's lightning detection challenges. Technological advances have contributed towards improving lightning detection and monitoring activities in many countries. South Africa has made considerably more progress in the field of lightning research than other African countries and possesses one of the three ground-based lightning detection networks in the southern hemisphere. However, despite these developments, rural communities in South Africa, and indeed in Africa, remain vulnerable to lightning, the occurrence of which is predicted to increase with climate change. A large proportion of the population of African countries resides in rural areas, where citizens participate in subsistence farming, and built infrastructure is not lightning safe. We recommend a call for the integration of indigenous and scientific knowledge as well as for the development of a participatory early warning system. Investigations into determining the most effective way to utilise existing monitoring networks – but with warning dissemination to rural communities – are also required. Lastly, future research on the development of lightning-safe rural dwellings or shelters, especially in lightning prone areas, is needed.

Significance:

- Climate change projections of increases in lightning incidence highlight an increased risk for vulnerable communities.
- There is a lack of literature focusing on lightning detection within rural communities.
- Technological advances now allow for better dissemination of lightning information and early warning within rural communities.
- The South African Lightning Detection Network is operational at a national level; however, there is no dissemination at a local level.
- There are currently no recommended design guidelines for informal dwellings and no safety protocols for rural communities.

Introduction

Lightning is one of the most frequently occurring geophysical phenomena.[1] Despite lightning being a familiar and researched phenomenon, it remains poorly understood, primarily due to the spontaneous spatial and temporal occurrence of lightning. There has been more than a century of research conducted on the physics and the phenomenology of lightning and yet some processes still require further in-depth research.[1,2] Lightning is complex and is usually accompanied by extreme weather such as hail, extreme wind gusts and heavy rainfall.[3,4] According to Blumenthal et al.[3], apart from incidental catastrophes and disasters, lightning strikes result in more deaths than any other natural event or phenomenon. In South Africa, the number of lightning deaths is about four times higher than the global average.[5] Although various studies report disparate lightning-related fatalities, the actual number may be higher as many injuries and deaths are often unreported.[6,7]

It is expected that extreme weather and the occurrence of lightning will also increase with climate change.[8] Africa has already experienced a warming trend, which is likely to continue in the future.[9] These climate change projections become an added concern for developing African countries that are already prone to lightning occurrences. Climate models support the positive correlation between lightning and global temperatures.[10] A study by Romps et al.[11] modelled the frequency of lightning strikes across the continental USA and predicted that lightning strike rates will increase significantly due to increases in global average air temperature. However, there is uncertainty regarding the expected changes in spatial distribution of lightning with climate change. Because climate change is intricately linked to almost all facets of society, developing countries are more likely to face the brunt of climate change due to their low adaptive capacity. Thus, monitoring and prediction of lightning incidences on a local scale for developing countries requires attention.

Lightning incidence itself may be a valuable variable that could be used to monitor climate change and severe weather changes.[7] There are now reliable ways for monitoring global and regional lightning activity in near real-time.[7] The ease of monitoring lightning across the globe using ground-based networks is frequently advocated. This ease makes lightning an attractive indicator for tracking changes in severe weather.[12] Climate models that include the impact of El Niño Southern Oscillation (ENSO) for South Africa predict increasing lightning activity as the climate becomes warmer and drier, despite the atmosphere becoming more stable.[10] These relationships indicate reduced rainfall, which may result in increased lightning incidence.[10]

Lightning has already been recognised as an important research topic on the African continent. Gijben[13] presents a review of the historical and current instrumentation used for the detection of lightning activity over South Africa. However, with an increase in the accessibility of detection methods and systems, it is important to re-evaluate their application in the South African context. Consequently, the objective of this review is to highlight the advances in lightning monitoring and detection, along with the major lightning detection challenges facing South Africa, including the relevance to rural communities.

Approach

In this review, we assess the latest lightning detection techniques and technologies used globally and include an update on the progress and challenges in lightning detection, with a focus on South Africa. A detailed summary of South Africa's current lightning research initiatives as well as future endeavours towards improving South Africa's lightning detection at a local/community level is highlighted. A mixed-methods approach was used, which included qualitative and quantitative inputs towards exploring existing lightning detection methods and investigating South Africa's current state of lightning detection. The focus was on the specific challenges facing South Africa, with respect to community-level lightning detection to provide feasible recommendations.

History of lightning detection

History internationally

Lightning was associated with God's anger up until the Middle Ages when the natural interpretation of attributing lightning to collisions between clouds by René Descartes began in the 17th century.[14] In 1746, Benjamin Franklin showed that lightning was electrical.[1] Franklin's well-known kite experiment in 1752 was a critical breakthrough in scientific research that showed that lightning was electrical.[1,2,6]

In the late 19th century, photography and spectroscopy became available as diagnostic techniques utilised for lightning research in England, Germany and the USA.[1,13] Investigations used time-resolved photographic techniques to identify individual 'strokes' consisting of a lightning discharge to the ground and the 'leader' that precedes the first strokes. In 1900, the double-lens streak camera was invented in England by Boys[15]. In the 1930s and after, Boys' double-lens camera was used in South Africa to study cloud-to-ground (CG) lightning.[16] Pockels[17-19] in Germany made the first lightning current measurements. Pockels analysed the residual magnetic field induced in basalt rock by nearby lightning currents and was able to estimate the values of those currents. Studies by Boys[15], Schonland[16], Pockels[17] and Uman[20] further elaborated on lightning photography and spectroscopy, while the early history of lightning photography and spectroscopy was comprehensively reviewed by Uman[21].

The modern era of lightning research dates back to work in England by Wilson[22], who investigated remote, ground-based electric field measurements. It was only about 20 years ago that transient luminous events and high-energy phenomena (runaway electrons, X-rays, and gamma rays including the terrestrial gamma ray flashes observed on orbiting satellites) were discovered and are still the subject of intensive present-day research.[1]

History in South Africa

In South Africa, lightning-related research can be traced back to the 1920s when Schonland and Malan, founding members of the Council for Scientific and Industrial Research (CSIR), pioneered the first electric field measurements in South Africa.[3] Much of the lightning research in South Africa was then continued and produced by Schonland and others during the 1930s.[23] The CSIR has continued to maintain its lightning research activities, and from the 1960s has actively participated in the development and testing of lightning detection equipment through the National Electrical Engineering Research Institute (NEERI) in Pretoria, and in cooperation with CIGRÉ (International Council on Large Electric Systems).[3]

In recent years, prominent South African institutions, including the South African Weather Service (SAWS), the University of the Witwatersrand and the University of KwaZulu-Natal have made significant contributions to the field of lightning research. The University of the Witwatersrand has led research on lightning medicine (keraunomedicine)[3] and lightning myths[24], whereas the nowcasting and forecasting of lightning threats[5] as well as the use of lightning to track the development of thunderstorms[25] in the country have been documented by the SAWS. Gijben et al.[5] recently developed a new lightning threat index for South Africa by using numerical weather prediction to enable forecasts of lightning threats. A study by Clulow et al.[26] conducted at the University of KwaZulu-Natal illustrated the use of ground-based lightning early warning systems for areas not covered by continent-wide lightning locating systems. These recent research studies and activities show the relevance of lightning through the ongoing advances in lightning research in South Africa.

Existing lightning detection techniques and systems

Existing lightning detection systems vary in terms of their spatio-temporal characteristics and identifying a suitable system for an application can therefore be complex (Figure 1; Table 1). Detection systems have different capabilities in terms of warning dissemination. Handheld detection systems for example have no dissemination capabilities and are spatially restricted, while national network systems have been integrated with global warning systems (i.e. the World-Wide Lightning Detection Network), and cover large areas (Figure 1).

Radiation that is emitted from lightning forms the basis of lightning detection and lightning location. During the lightning process, electromagnetic and acoustic radiation is generated in various forms, which include radio emission (occurs in the form of short pulses), optical radiation (emitted by thermal radiation of the hot lightning channel) and acoustical radiation (mainly human based).[6,28]

During the last 50 years, various lightning mapping systems have been developed and operate at various frequency ranges and bandwidths. Ground-based detection systems using multiple antennas, space/satellite-based systems and mobile systems using a direction and a sense antenna in the same location are currently the three common types of lightning detection systems globally.[28] The most commonly used techniques remain the ground- and space-based lightning detection networks.[29] These networks are continuously improving, and their data are growing in importance for scientists and operational weather forecasters.

Satellite/space-based systems

Tracking thunderstorms and assessing cyclone intensification become important challenges in weather prediction for remote regions where surface observations and radar systems are not available. Significant advances in the understanding of global distribution and frequency of lightning have been made possible by the different types of satellite-borne lightning detectors.[30] Two primary satellite sensors that have been widely used include the National Aeronautics and Space Administration (NASA) Optical Transient Detector (OTD) and the Lightning Imaging Sensor (LIS) on board the Tropical Rainfall Measuring Mission (TRMM) satellite.[10,28,31] The OTD and the LIS are both low Earth orbit instruments, capable of detecting optical pulses from lightning flashes, during both day and night. However, they do not accurately separate CG and cloud lightning incidence. Additionally, the majority of these satellites, such as the LIS, are polar-orbiting satellites, with limited spatio-temporal coverage. They are also incapable of providing near real-time lightning monitoring, detection and warning.[30]

The OTD was operational from May 1995 until March 2000 with a spatial resolution of 8 km, while LIS, with a 4-km spatial resolution, was active from November 1997 to 2015.[32] The OTD and LIS systems were critically assessed, providing guidance on the applicability for research use and instruction for new instrument design.[31-35] A merged global lightning 0.5°-resolution data set composed of the individual LIS and OTD orbits is freely available online at http://lightning.nsstc.nasa.gov/data/data_lis-otd-climatology.html.[36] The LIS-OTD climatology is the most accurate depiction of total lightning across the planet to date and is named the High-Resolution Annual Climatology database.[37]

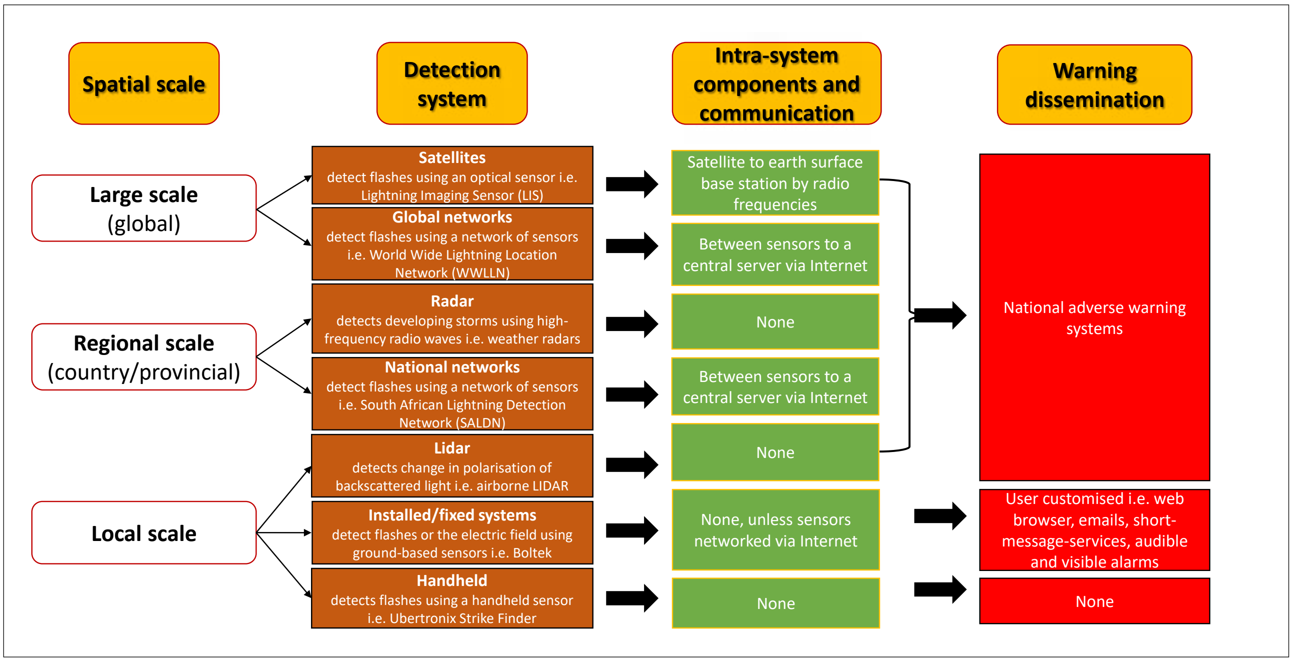


Figure 1: The spatial characteristics of existing lightning detection systems and their associated warning and alert dissemination capabilities.

Table 1: A list of present-day lightning detection options

Present-day lightning detectors	Generalised description
Radio frequency (RF) detectors	Measure past energy discharges from lightning and can determine the approximate distance and direction of the threat.
Inferometers	Multi-station devices, more costly than RF detectors, employed for research purposes and require a skilled operator.
Network systems	Multiple ground-based RF sensors are networked to determine location of lightning over large spatial scales (i.e. continent or country scale).
Electric field meters (EFM)	A pre-lightning sensor that measures the potential gradient (voltage) changes of the Earth's electric field (cloud voltages) and reports changes from predetermined thresholds to lightning breakdown values. They consist of a narrow reporting range and false alarms may occur from various sources such as dust storms.
Optical monitors	Detect light flashes from cloud-to-cloud lightning that typically precedes cloud-to-ground lightning.
Hybrid designs	Consists of a combination of the aforementioned technology designs such as RF and EFM sensors.
Meteorological information services	Includes meteorological subscription services usually sourced from a network system.

Source: Kithil[27]

The new generation series of GOES-R (Geostationary Operational Environmental Satellites) carrying a Geostationary Lightning Mapper (GLM) was launched in November 2016 and has been deployed in a geostationary orbit to continuously detect lightning activity over America and its adjacent ocean region in the western hemisphere.[38] The GLM is an optical sensor that detects total lightning (in-cloud, cloud-to-cloud and cloud-to-ground) activity over the western hemisphere.[39] The GLM delivers lightning measurements similar to those of LIS but provides continuous lightning detection.[38] This GLM will be able to provide high-quality data for forecasting severe storms and convective weather but only over the western hemisphere.[38]

EUMETSAT plan to launch the Meteosat Third Generation (MTG) operational meteorological satellite in 2021 with an onboard lightning imager operating on a continuous basis covering the entire MTG disk (including the entire African continent). It is expected to deliver near real-time information on total lightning for the purpose of supporting nowcasting of severe weather warnings and monitoring deep convection.[30] The development of such modern lightning detection instrumentation has been driven by a variety of practical needs and applications as well as research needs.[40]

Ground-based lightning detection networks

Ground-based global lightning observation networks are based on the Schumann resonance method.[10,41] As a result of the narrow time scale (sub-millisecond to millisecond) and the large spatial scales associated with the lightning current, the majority of the energy in the radiated spectrum is contained in the extremely low frequency (ELF 3–3000 Hz) and very low frequency (VLF 3–30 kHz) bands.[4] Further details on these different frequency ranges are given by Cummins et al.[42]

Electromagnetic waves disseminate at ELF and VLF frequencies, by being reflected from the ground and from the conducting layer of the atmosphere known as the ionosphere, and in this manner, they can travel large distances around the Earth.[4] The low loss propagation of sferics (typically 2–3 dB/1000 km) allow measurements to be conducted spatially from their source locations within the ionosphere waveguide. This makes networks of ELF/VLF sensors particularly useful in long-range severe weather monitoring applications, compared to weather radars. Weather radars use microwave-frequency radar beams that are blocked by mountainous regions when locating the presence of storms over several kilometres.[4]

According to Mayekar et al.[4], a low-frequency lightning receiver forms part of a node/sensor that captures electromagnetic radiation emitted by lightning. Several such nodes/sensors are distributed across a certain geographical area (to form a network), which rely on the use of either time-of-arrival[28,40,42] or magnetic direction-finding[28,40] techniques to detect lightning. The digitised data are sent to a central processing system, which processes these data to calculate the lightning signal characteristics such as peak current, polarity and source location. Finally, the central processing system sends this information to a user/display software.

Examples of regional LDNs include the South African Lightning Detection Network (SALDN) in South Africa[13], the European Cooperation for Lightning Detection (EUCLID)[28] and the National Lightning Detection Network (NLDN) in the USA[40,43]. Regional LDNs operate with sensors spaced relatively close to each other (e.g. the SALDN consists of 24 sensors spread across South Africa)[5], providing regional coverage of total lightning with high detection efficiency. These networks, however, do not provide information over oceanic regions or remote locations where no sensors are installed.[38] Over the years, long-range detection networks have also been developed to enable global coverage and real-time lightning detection, but with a lower detection efficiency than short-range detection networks and satellite detection systems.[38] Global LDNs consist of sensors separated by thousands of kilometres. Examples of global LDNs include the Global Lightning Dataset (GLD360)[44] and the World Wide Lightning Location Network (WWLN).[45] These sensors detect mainly CG with regional LDNs also detecting a small fraction of cloud-to-itself lightning.[5]

Recently, an innovative lightning location network, 'Blitzortung Lightning', has been established. This network is a worldwide, real-time, community collaborative network, and has been available since 2003. The network monitors magnetic field (H-field) and electrical field (E-field) emissions from lightning strikes and has a set of servers in Europe to correlate the time-of-arrival at detectors vs GPS-time to locate strikes.[46] A real-time, online map is available which displays strike information for North America. The web application notifies users via email, SMS (short message service) or URL-call when lightning is detected within their area. Strikes are colour coded to show how recent they are. Currently, coverage is biased towards the largest clusters of lightning detectors across Europe, USA and Australia, whilst Africa, Asia and South America remain devoid of detectors.

Handheld/mobile

Handheld lightning detectors allow users the opportunity to buy a detector easily from a retail store and set it up themselves instead of having to pay for a service or for lightning information. The cost of these devices varies according to the accuracy and design of the equipment. Such lightning detection instrumentation typically has limitations and the value of these portable devices requires consideration. They detect mostly the intensity of the electromagnetic pulse (EMP)[47] and are generally unable to detect cloud-to-cloud lightning (which usually precedes CG strikes), which is critical in recognising an approaching storm. Additional limitations include, but are not limited to, poor detection ranges, inability to determine direction or location of a lightning strike as well as interferences from other EMP-emitting devices (such as electrical equipment, fluorescent lights, appliances and even car engines), which may result in either missed strikes or false alarms. Examples of popular hand-held devices include the Ubertronix Strike Finder (Ubertronix, Inc. San Antonio, Texas, USA), designed to record lightning strikes during the day and night by using an infrared sensor and microcontroller-based technology,[48] while other options include the ThunderBolt Storm Detector (Storm Systems, Tampa, Florida, USA), SkyScan lightning detector (Extreme Research Corporation, Port Richey, Florida, USA) and the INO Weather Pro portable weather station (INO Technologies, Louisville, Colorado, USA).

Lightning detection in South Africa

Current lightning detection system

The SAWS had no role in measuring lightning activity prior to 2005.[49] Eskom, the major power utility of South Africa, operated a network of six Lightning Position and Tracking System lightning detection sensors.[50] Before this, the CSIR operated a lightning detection network of 400 lightning flash counters and was the first institution to produce a lightning flash density map for South Africa.[49]

Over recent years, the detection of lightning occurrences in South Africa has been undertaken by the SALDN, which is operated by the SAWS.[5] In 2005, the SAWS purchased a Vaisala, lightning detection network (LS 7000 and LS 7001, Helsinki, Finland), making South Africa one of only three countries in the southern hemisphere to operate such a network, with the others being in Brazil and Australia.[49] The network provided the SAWS with its first opportunity to explore lightning and also to provide lightning information to the public.[13] The network now consists of 24 sensors across the country.[5]

Lightning research initiatives

In recent years, the University of the Witwatersrand has participated in lightning research in South Africa and has been a key role player in the development of a multidisciplinary interest group called Lightning Interest Group for Health, Technology and Science (LIGHTS).[3] LIGHTS has been successfully running since 2015, contributing, disseminating and sharing vital information regarding lightning and lightning research in South Africa and within the broader African lightning community. The African Centres for Lightning and Electromagnetics Network (ACLENet) is a pan-African network of Centres that is dedicated to reducing infrastructure damage, injury and mortality resulting from lightning across Africa.[51] It operates as a not-for-profit and non-governmental organisation with national centres in Zambia, Malawi, Kenya and South Africa. The network consists of several research and technical advisors who are internationally recognised and serve voluntarily to advise ACLENet on education, research and grant proposals, mentor African researchers, supervise graduate studies and promote ACLENet worldwide. The ACLENet is designed to be user-friendly and can be translated online into Arabic, French, Portuguese, Spanish and Swahili. The network gathers and presents media articles about lightning injuries and deaths caused by lightning, which are listed by country. The Earthing and Lightning Protection Association (ELPA) is also a significant contributor towards the standard of safety in the South African lightning and protection industry. ELPA offers certification of qualified designers, installers and inspectors, with recognition by the University of the Witwatersrand, and the South African Institute of Electrical Engineers (SAIEE) amongst others.[52] LIGHTS, ACLENet and Eskom are among the collaborating institutions. ELPA has been established as a non-profit organisation of voluntary membership. The University of Zambia has also recently contributed towards the academic knowledge regarding lightning in Africa and has initiated an MSc and PhD programme in high voltage, electromagnetic compatibility, lightning studies and protection.[51]

In addition to the SALDN disseminating lightning warnings across South Africa via media broadcasts, a few other initiatives exist for alerting South Africans to possible threats from lightning. One such example is the WeatherBug application. WeatherBug is a mobile application brand owned by GroundTruth, a company based in New York City.[53] This mobile application provides near real-time lightning detection and provides alerts via the application. WeatherBug uses data from the Total Lightning Network (run by Earth Network) together with GPS location data from the users' mobile phones.[54] The Total Lightning Network dates to 2009, with initially most sensors existing in the USA. The network now covers areas of North and South America, Africa, Asia, Europe and Australia. AfricaWeather (Johannesburg, South Africa) is another example of a mobile application disseminating lightning warnings to South Africans. It is the only South African built application with lightning and storm detection capability.[55] The application provides basic free content (daily weather notices), while advanced features (including lightning proximity and lightning data) require a paid subscription. The online storm-tracking tool allows individuals to identify the location, intensity and time of recorded lightning strikes. AfricaWeather monitors the country's grounded lightning strikes using the Earth Network's (Germantown, Maryland, USA) Total Lightning Network.[55] Alerts are disseminated to numerous schools and golf courses across South Africa through a siren and spinning strobe light that is installed and maintained through a paying subscription, and to a list of specified contacts through SMSs.

Community-level lightning detection in South Africa: A new approach

Despite great strides being made in detecting lightning throughout the world, including South Africa, there remains a high number of lightning fatalities in many rural communities within developing countries such as South Africa. According to media articles and reports, several of these lightning fatalities occur whilst rural people are still present inside their homesteads. However, there is still a lack of literature that focuses on lightning detection within rural communities. Furthermore, no literature exists on determining effective approaches to communicate lightning data, threats and advance warnings in a manner appropriate for rural communities, as well as information on how to reduce lightning damage in rural dwellings. Such information is vital to assess risk knowledge as part of early warning systems, which is accounted for in the dissemination aspects to build response capabilities that will enable mitigation.

Based on the detection and warning systems reviewed, there are a number of community-level, automated possibilities available for South Africa that are appropriate for high-risk lightning areas. The first is that the national lightning detection network (SALDN) supplies areas with lightning warnings using the method employed by AfricaWeather. The second option includes a local measurement system, consisting of a single sensor/node. Numerous types of local measurement systems exist and continue to evolve. Examples of these stand-alone systems include the Boltek lightning detection systems (Port Colborne, Ontario, Canada) and the Campbell Scientific lightning warning systems (Logan, Utah, USA). These systems are not only capable of detecting lightning strikes, but are also capable of monitoring the electric field changes by using an electric field meter and providing warnings before the first lightning strike takes place. These systems are, however, expensive, but there are more cost-effective lightning flash sensors now included with some basic weather stations. The ATMOS41 by the METER Group (Pullman, Washington, USA) features a lightning strike counter with distance categories, as well as other meteorological sensors. A third approach includes identifying lightning prone communities using the SALDN and installing 'sacrificial towers' containing a lightning rod to divert the lightning pathway from dwellings.

Currently, in South Africa, the SALDN operates at a national level and has the capability of disseminating lightning data to a local level. However, this dissemination has not been implemented and rural communities continue to lack cognisance of the dangers of lightning. This remains a significant gap within lightning detection research in South Africa and a dire need exists to bridge the gap between the SALDN and rural communities.

Lightning detection challenges in South Africa

People residing in South Africa's rural areas are often outdoors due to work activities such as subsistence farming and livestock herding. Such individuals are the most prone to facing lightning-related risks.[24] The houses in rural communities are commonly not well earthed and provide little protection against lightning. Consequently, some lightning deaths occur whilst people are inside their homes. Many rural structures do not contain metal plumbing, electrical wiring or reinforcing steel that provide a pathway for a lightning current to be grounded.[24] Rural dwellings also do not have proper interior flooring, which increases the risk, as many deaths are due to ground currents from nearby lightning strikes, rather than direct strikes. Furthermore, rural housing often has thatched roofs or newspaper to insulate the roof, both of which are a fire risk.[24] Rural areas therefore lack lightning-safe shelters, and also have fewer fully enclosed metal-topped vehicles, leaving communities vulnerable to the threat of lightning. The economic implications and feasibility of building lightning-safe houses and structures as well as the installation of lightning detectors in high lightning risk areas requires further investigation. Investigation specifically into developing lightning-safe shelters is the key priority, and includes projects to fund and develop lightning safe shelters (such as schools, community halls, as well as lightning safe houses). This is urgently needed in South Africa where funders to support such initiatives are required. There also appears to be no design criteria for establishing lightning-safe rural dwellings which is a critical need in South Africa.

Several cultural beliefs have been associated with lightning strikes (and thunder) in South Africa, which include mythical association. Indigenous South Africans have religious and traditional beliefs that lightning may be directed to strike someone, and that significant personality changes ensue after a lightning strike, and that it could be a sign of God's anger.[3,24,56] Such myths still exist and some hinder the necessity to take precautionary/mitigation measures, thereby increasing the risk of lightning injury.[56] This situation calls for the integration of selected relevant indigenous and scientific knowledge into educational packages that are relevant to rural community inhabitants.

The review of techniques shows that various lightning detection systems have evolved over recent years; however, the warnings are not disseminated well to rural communities. Various practical constraints such as poor network signals, a lack of knowledge and the cost of smartphones and data, are prohibitive to the success of such lightning detection systems. Consequently, rural communities continue to remain vulnerable to lightning threats.

Way forward and recommendations

Significant progress on South Africa's national lightning detection and monitoring has been achieved. Despite these advances, local level research and, more specifically, the vulnerability of rural communities to lightning incidence and threats require further attention as rural communities continue to live without any lightning warning. The proposed way forward for improving lightning detection on a local scale is through a system with monitoring and predictive capacity to improve the detection of lightning occurrences and assist rural communities in preparing for lightning through risk knowledge and near real-time/early warning systems is ultimately needed. This can be achieved through the communication and dissemination of alerts in a timeous and comprehensible manner in languages that are understood within specific communities. Building lightning-safe rural dwellings and shelters is also required as well as transformative adaptation. It has been shown that in rural areas using participatory research methodologies, as well as community-based adaptation planning, that adaptation can become an iterative co-learning process and facilitate transformative adaptation through the integration of indigenous knowledge with science-based systems. An opportunity therefore exists for bridging the gap between the existing SALDN and rural communities.

To raise awareness of lightning, a national lightning awareness week should be introduced to coincide with that run internationally to promote the magnitude of risks associated with lightning and how to minimise risks, especially in rural communities.

Conclusions

The current study provides a synthesis on the development and detection of lightning activity internationally and at a local level (South Africa). There are different lightning detection systems available, which vary in their spatial scale, detection and dissemination capability. The literature revealed significant and ongoing advances in detection methods, mainly using satellites, but the vulnerability of rural communities in countries like South Africa remains a challenge, mainly due to the insufficient dissemination of lightning warnings. The SALDN continues to accurately detect lightning activity at a national level but warnings are not disseminated to a local scale. Also, even if warnings are disseminated at a local scale, there are few lightning safe shelters/dwellings or fully enclosed metal-topped motor vehicles available in rural areas.

Myths and beliefs regarding lightning in rural areas also continue to remain a challenge in South Africa and hinder necessary precautionary measures. The national school education system needs to include lightning safety and the role of cultural beliefs associated with lightning. Education around lightning safety, the development of lightning safety protocols/guidelines and the involvement of multiple stakeholders – from community members and government extension officers to non-governmental organisations – is required.[56]

Acknowledgements

We acknowledge the financial assistance of the Durban Research Action Partnership (DRAP) as well as the uMngeni Resilience Project, funded by the Adaptation Fund, which is a partnership between the South African Department of Environment, Forestry and Fisheries, the South African National Biodiversity Institute, the uMgungundlovu District Municipality and the University of KwaZulu-Natal's Centre for Transformative Agricultural and Food Systems towards this research. Opinions expressed and conclusions arrived at, are those of the authors and are not necessarily to be attributed to the funders.

Competing interests

We declare that there are no competing interests.

Authors' contributions

M.M. was responsible for writing the initial draft. A.D.C. assisted with technical aspects, S.S. and M.J.S. with theoretical aspects and T.M. with community resilience aspects. A.D.C., S.S., M.J.S. and T.M. also provided supervision, assisted with reviewing and provided editorial contributions.

References

1. Dwyer JR, Uman MA. The physics of lightning. Phys Rep. 2014;534(4):147–241. http://dx.doi.org/10.1016/j.physrep.2013.09.004
2. Akinyemi M, Boyo A, Emetere M, Usikalu M, Olawole F. Lightning a fundamental of atmospheric electricity. IERI Procedia. 2014;9:47–52. https://doi.org/10.1016/j.ieri.2014.09.039
3. Blumenthal R, Trengrove E, Jandrell IR, Saayman G. Lightning medicine in South Africa. S Afr Med J. 2012;102(7):625–626. https://doi.org/10.7196/samj.5219
4. Mayekar M, Kulkarni A, Marakarkandy B. Designing and simulation of low frequency cloud to ground lightning receiver for severe weather monitoring application. Procedia Comput Sci. 2015;45:661–670. https://doi.org/10.1016/j.procs.2015.03.128
5. Gijben M, Dyson LL, Loots MT. A lightning threat index for South Africa using numerical weather prediction data [dissertation]. Pretoria: University of Pretoria; 2016.
6. Bhavika B. The influence of terrain elevation on lightning density in South Africa [thesis]. Johannesburg: University of Johannesburg; 2007.
7. Frisbie P, Colton J, Pringle J, Daniels J, Meyers M. A forecasting methodology that uses moisture parameters to pinpoint locations of potential lightning. Central Region Technical Attachment. 2013, Number 13-01.
8. Ziervogel G, New M, Archer van Garderen E, Midgley G, Taylor A, Hamann R, et al. Climate change impacts and adaptation in South Africa. Clim Change. 2014;5(5):605–620. https://doi.org/10.1002/wcc.295
9. Hulme M, Doherty R, Ngara T, New M, Lister D. African climate change: 1900-2100. Clim Res. 2001;17(2):145–168. https://doi.org/10.3354/cr017145
10. Price C. Thunderstorms, lightning and climate change. In: Betz HD, Schumann U, Laroche P, editors. Lightning: Principles, instruments and applications. Dordrecht: Springer; 2009. p. 521–535. https://doi.org/10.1007/978-1-4020-9079-0_24
11. Romps DM, Seeley JT, Vollaro D, Molinari J. Projected increase in lightning strikes in the United States due to global warming. Science. 2014;346(6211):851–854. https://doi.org/10.1126/science.1259100
12. Price CG. Lightning applications in weather and climate research. Surv Geophys. 2013;34(6):755–767. https://doi.org/10.1007/s10712-012-9218-7
13. Gijben M. The lightning climatology of South Africa. S Afr J Sci. 2012;108(3–4):44–53. https://doi.org/10.4102/sajs.v108i3/4.740
14. Kasparian J, Wolf JP. On lightning control using lasers. In: Yamanouchi K, Giulietti A, Ledingham K, editors. Progress in ultrafast intense laser science. Springer Series in Chemical Physics. Berlin, Heidelberg: Springer; 2010. https://doi.org/10.1007/978-3-642-03825-9_6
15. Boys CV. Progressive lightning. Nature. 1926;118:749–750. https://doi.org/10.1038/118749a0
16. Schonland B. The lightning discharge. Handb Phys. 1956;22:576–628.
17. Pockels F. Ueber das magnetische Verhalten einiger basaltischer Gesteine [On the magnetic behaviour of some basaltic rocks]. Ann Phys. 1897;299(13):195–201. German. https://doi.org/10.1002/andp.18972991329
18. Pockels F. Bestimmung maximaler Entladungsstromstärken aus ihrer magnetisirenden Wirkung [Determination of maximum discharge currents from their magnetising effect]. Ann Phys. 1898;301(6):458–475. German. https://doi.org/10.1002/andp.18983010620
19. Pockels F. Über die Blitzentladungen erreichte Stromstärke [Amperage reached via the lightning discharges]. Ann Phys. 1900;Z2:306–307. German.
20. Uman MA. Lightning. New York: Courier Corporation; 2012.
21. Uman MA. Lightning. New York: McGraw-Hill; 1969.
22. Wilson C. Investigations on lightning discharges and on the electric field of thunderstorm. Philos Trans R Soc Lond. A. 1921;221:73–115. https://doi.org/10.1098/rsta.1921.0003
23. Rakov VA, Uman MA. Lightning: physics and effects. New York: Cambridge University Press; 2003.
24. Trengove E. Lightning myths and beliefs in South Africa: Their effect on personal safety. Nat Hazards. 2015;77(1):101–110. https://doi.org/10.1007/s11069-014-1579-4
25. Gijben M, De Coning E. Using satellite and lightning data to track rapidly developing thunderstorms in data sparse regions. Atmosphere. 2017;8(40):67. https://doi.org/10.3390/atmos8040067
26. Clulow AD, Strydom S, Grant B, Savage MJ, Everson CS. Integration of a ground-based lightning warning system into a mining operation in the Democratic Republic of the Congo. Weather Clim Soc. 2018;10(4):899–912. https://doi.org/10.1175/WCAS-D-18-0004.1
27. Kithill R. Overview of lightning detection equipment. Louisville, CO: US National Lightning Safety Institute; 2012. Available from: http://lightningsafety.com/nlsi_lhm/detectors.html
28. Finke U, Kreyer O. Detect and locate lightning events from geostationary satellite observations. Technical Report Part I EUM/CO/02/1016/SAT: Review of existing lightning location systems. Hannover: Institute für Meteorologie und Klimatologie, Universität Hannover; 2002.
29. Rudlosky SD, Fuelberg HE. Documenting storm severity in the mid-Atlantic region using lightning and radar information. Mon Weather Rev. 2013;141(9):3186–3202. https://doi.org/10.1175/mwr-d-12-00287.1
30. Labrador L. The detection of lightning from space. J Atmos Sci. 2017;72(2):54–59. https://doi.org/10.1002/wea.2903
31. Christian HJ, Blakeslee RJ, Boccippio DJ, Boeck WL, Buechler DE, Driscoll KT, et al. Global frequency and distribution of lightning as observed from space by the Optical Transient Detector. J Geophys Res Atmos. 2003;108(D1):ACL4-1–ACL4-15. https://doi.org/10.1029/2002jd002347
32. De Boer AM, Collier AB, Caballero R. Processes driving thunderstorms over the Agulhas Current. J Geophys Res Atmos. 2013;118(5):2220–2228. https://doi.org/10.1002/jgrd.50238
33. Boccippio DJ, Koshak WJ, Blakeslee RJ. Performance assessment of the optical transient detector and lightning imaging sensor. Part I: Predicted diurnal variability. JTECH. 2002;19(9):1318–1332. https://doi.org/10.1175/1520-0426(2002)019<1318:paotot>2.0.co;2
34. Nickolaenko A, Hayakawa M, Sekiguchi M. Variations in global thunderstorm activity inferred from the OTD records. Geophys Res Lett. 2006;33(6), Art. #L06823. https://doi.org/10.1029/2005gl024884
35. Xu W, Zipser EJ, Liu C, Jiang H. On the relationships between lightning frequency and thundercloud parameters of regional precipitation systems. J Geophys Res Atmos. 2010;115(D12), Art. #D12203. https://doi.org/10.1029/2009jd013385
36. Cecil, DJ, Buechler DE, Blakeslee RJ. Gridded lightning climatology from TRMM-LIS and OTD: Dataset description. Atmos Res. 2014;135:404–414. https://doi.org/10.1016/j.atmosres.2012.06.028
37. Albrecht RI, Goodman SJ, Buechler DE, Blakeslee RJ, Christian HJ. Where are the lightning hotspots on Earth? Bull Amer Meteor Soc. 2016;97(11):2051–2068. https://doi.org/10.1175/bams-d-14-00193.1

38. Bürgesser RE. Assessment of the World Wide Lightning Location Network (WWLLN) detection efficiency by comparison to the Lightning Imaging Sensor (LIS). Q J Royal Meteorol Soc. 2017;143(708):2809–2817. https://doi.org/10.1002/qj.3129

39. GOES-R. Geostationary Operational Environmental Satellite -R Series. Greenbelt: GOES-R Series Program; 2018. Available from: https://www.goes-r.gov/

40. Cummins KL, Murphy MJ. An overview of lightning locating systems: History, techniques, and data uses, with an in-depth look at the US NLDN. IEEE Electromagnetic Compatibility. 2009;51(3):499–518. https://doi.org/10.1109/temc.2009.2023450

41. Price C, Pechony O, Greenberg E. Schumann resonances in lightning research. J Light Res. 2007;1:1–15.

42. Cummins KL, Murphy MJ, Tuel JV. Lightning detection methods and meteorological applications. Paper presented at: IV International Symposium on Military Meteorology; 2000 September 26–28; Malbork, Poland.

43. Orville RE, Huffines GR, Burrows WR, Cummins KL. The North American lightning detection network (NALDN) – Analysis of flash data: 2001–09. Mon Weather Rev. 2011;139(5):1305–1322. https://doi.org/10.1175/2010mwr3452.1

44. Mallick S, Rakov V, Ngin T, Gamerota W, Pilkey J, Hill J, et al. Evaluation of the GLD360 performance characteristics using rocket-and-wire triggered lightning data. Geophys Res Lett. 2014;41(10):3636–3642. https://doi.org/10.1002/2014gl059920

45. Virts KS, Wallace JM, Hutchins ML, Holzworth RH. Highlights of a new ground-based, hourly global lightning climatology. Bull Amer Meteor Soc. 2013;94(9):1381–1391. https://doi.org/10.1175/bams-d-12-00082.1

46. Saratoga-Weather.org [homepage on the Internet]. No date [cited 2014 Jul 01]. Available from: saratoga-weather.org

47. Ambient Weather. Personal lightning detector purchase guide [webpage on the Internet]. c2019 [cited 2019 Jun 25]. Available from: https://www.ambientweather.com

48. Strike Finder™ [homepage on the Internet]. c2019 [cited 2019 Jun 25]. Available from: https://strikefinder.com

49. Gill T. A lightning climatology of South Africa for the first two years of operation of the South African Weather Service Lightning Detection Network: 2006–2007. Paper presented at: 20th International Lightning Detection Conference; 2008 April 21–23; Vaisala, Tucson, Arizona, USA.

50. Ngqungqa S. A critical evaluation and analysis of methods of determining the number of times that lightning will strike a structure [dissertation]. Johannesburg: University of the Witwatersrand; 2005.

51. ACLENet. ACLENet, lightning and electromagnetics network. Kampala: Firespring; 2019. Available from: https://aclenet.org/

52. ELPA. Our 7 core principles [webpage on the Internet]. c2019 [cited 2019 Jan 24]. Available from: https://elpasa.org.za/

53. Earth Networks. WeatherBug [webpage on the Internet]. No date [cited 2019 Jan 24]. Available from: https://www.earthnetworks.com/weatherbug/

54. GroundTruth. WeatherBug [homepage on the Internet]. c2018 [cited 2019 Jan 24]. Available from: https://www.weatherbug.com

55. AfricaWeather. Golf courses and schools [webpage on the Internet]. c2019 [cited 2019 Mar 09]. Available from: https://www.africaweather.com/products/golfschools

56. Dlamini WM. Lightning fatalities in Swaziland: 2000–2007. Nat Hazards. 2009;50(1):179–191. https://doi.org/10.1007/s11069-008-9331-6

The altitude of sprites observed over South Africa

AUTHORS:
Dakalo C. Mashao[1,2]
Michael J. Kosch[2,3,4]
Jozsef Bór[5]
Stanislaus Nnadih[6,7]

AFFILIATIONS:
[1]Department of Physics, University of KwaZulu-Natal, Durban, South Africa
[2]Space Science Division, South African National Space Agency, Hermanus, South Africa
[3]Department of Physics, University of the Western Cape, Cape Town, South Africa
[4]Physics Department, Lancaster University, Lancaster, United Kingdom
[5]Research Centre for Astronomy and Earth Science, Hungarian Academy of Science, Budapest, Hungary
[6]African Regional Centre for Space Science and Technology Education, Ile Ife, Nigeria
[7]SpaceLab, Department of Electrical Engineering, University of Cape Town, Cape Town, South Africa

CORRESPONDENCE TO:
Dakalo Mashao

EMAIL:
mashaodakalo@gmail.com

DATES:
Received: 07 Feb. 2020
Revised: 12 June 2020
Accepted: 13 June 2020
Published: 29 Jan. 2021

HOW TO CITE:
Mashao DC, Kosch MJ, Bór J, Nnadih S. The altitude of sprites observed over South Africa. S Afr J Sci. 2021;117(1/2), Art. #7941. https://doi.org/10.17159/sajs.2021/7941

ARTICLE INCLUDES:
☒ Peer review
☒ Supplementary material

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

EDITOR:
Amanda Weltman

KEYWORDS:
charge moment change, lightning, transient luminous events

FUNDING:
Hungarian National Research, Development and Innovation Office (K115836), South African National Space Agency, University of the Western Cape

Sprites are mesospheric optical emissions that are mostly produced by large, positive cloud-to-ground lightning discharges. Sprites appear in different morphologies such as carrot, jellyfish and column, and are typically in the altitude range of ~40–100 km above the Earth's surface. Sprites are a subset of transient luminous events and they contribute to the global electric circuit. South Africa has large convective thunderstorms, which typically occur in the summer months of every year. Peak current, time and geographical position of lightning strokes were obtained from the South African Weather Service. Sprite observations were recorded in South Africa for the first time on 11 January 2016 from Sutherland in the Northern Cape using a night-vision television camera from the South African National Space Agency's Optical Space Research laboratory. We report the first estimates of the top altitude, and the altitude of maximum brightness, of 48 sprites over South Africa. We found that the average top altitude and the altitude of maximum brightness of sprites are approximately 84.3 km and 69 km, respectively, which is consistent with estimates made elsewhere. We also found a moderately high positive and a weak positive correlation between the top altitude and the altitude of maximum brightness, respectively, of sprites and the lightning stroke charge moment change.

Significance:

- We present the first altitude estimation of sprites observed over Africa.
- The altitude of sprites observed over South Africa is in agreement with observations made elsewhere.
- There is a positive correlation between the top altitude of sprites and the parent lightning charge moment change.
- Sprite maximum brightness is observed near the stratopause.

Introduction

Sprites are optical phenomena generated by the electric field in the mesosphere, almost exclusively during positive cloud-to-ground (+CG) lightning flashes with a high peak current and a high charge moment change (CMC). Sprites play a role in the global electric circuit.[1-3] The optical signature of sprite events typically lasts for about 1–10 ms.[2] However, some sprite events can last for more than 10 ms, especially in the case of dancing sprites and column sprites which are not very bright.[4,5] Sprites are the most familiar type of transient luminous events (TLEs), which are short-lived gas breakdown phenomena that occur well above the altitudes of normal lightning and storm clouds, typically ~40–100 km. Sprites, halos, elves, blue jets and gigantic jets are different types of TLEs. Sprites consist of many streamers, which are typically measure 60–200 m in horizontal length.[1]

Sprites are generated by the quasi-electrostatic field generated by the large CMC from the initial lightning stroke and its subsequent continuing current during mesoscale convective system thunderstorms. This occurs when a high amount of positive charge is lowered to the ground during a lightning discharge and the opposite sign charge above the convective thundercloud creates a quasi-electrostatic field in the mesosphere. This results in heating and ionisation in the mesosphere, leading to the initiation of a streamer propagation which is observed as sprites.[1,6,7] A CMC of about 120–18 600 C km with a discharge time of approximately hundreds of microseconds has been found in the parent lightning strokes of sprites.[1,6,8,9] A CMC value of 1000 C km or more has a 90% chance of generating TLEs, whereas a CMC value of 600 C km or less has only a 10% chance of producing TLEs.[6]

The first reported sprite was recorded by chance on 6 July 1989 in the USA.[10] Since then, TLE-related research has been active in many continents, for instance, the Americas, Africa, Asia and Europe.[1,2,4-12] Sprite events have been recorded from the ground and space (e.g. by the ISUAL sensor).[13] There have been several reports on the altitude of sprites. Sentman et al.[14] conducted the first sprites triangulation. They found that sprites' average top altitude was about 88 km, with a root mean square error of approximately 5 km. The top altitude for the largest and brightest event that they recorded occurred at about 95 km. Wescott et al.[15] investigated column sprites or c-sprites. They found that the top altitude for these sprites varied from 81.3 km to 88.9 km, and the mean value of the top of the column sprites was 86.4 km, with a standard deviation of 1.9 km.

Stenbaek-Nielsen et al.[16] found that the triangulated initiation altitude of the downward streamer of sprites ranged from 66 km to 89 km and the triangulated top of the sprites varied from 79 km to 96 km. Stenbaek-Nielsen et al.[16] also analysed the sprites data which they observed from one location using high-speed cameras. They assumed that the sprite occurred close to the causal lightning strike and found that onset altitude ranged from 75 km to 95 km above the Earth's surface, with an average altitude of 85 km. The initiation altitude uncertainty of the sprite was determined to be about 1 km, depending on the observation elevation angle. Ground-based sprite observations were conducted on the eastern coastline of the Mediterranean Sea for the first time in the winter of 2005/2006 by Ganot et al.[17] They found that the average altitude of the bright transition region of sprites, measured between the diffuse region and the streamer region, was 73±10 km above the Earth's surface.[17]

Li et al.[18] reported the lightning CMC associated with long-delayed sprites. From their high camera frame rate time-resolution studies, they found that short delay (<15 ms) sprites initiate at an altitude of about 75–80 km.

A lightning CMC of 600 C km is required to initiate a minimum long-delayed sprite (>15 ms). The very long delay (>120 ms) sprites are initiated at an altitude of approximately 70 km above the Earth's surface and are associated with a minimum lightning CMC of 2000 C km. McHarg et al.[19] investigated sprites' streamers at 10 000 frames per second (fps) with a video frame exposure time of 50 μs. They found that streamers initiate at an altitude of about 82 km and propagate downward to about 45 km. The upward streamers start on the sprite body at an altitude near 75 km and propagate up to 95 km.[19]

Sprites were recorded for the first time in South Africa on 11 January 2016 by Nnadih et al.[12] This work is aimed at characterising the altitude of these sprites observed over South Africa.

Observations and measurements

All the sprites reported here were recorded from the South African Astronomical Observatory (SAAO) in Sutherland, Northern Cape, South Africa (20.81°E, 32.39°S). SAAO is located at an altitude of ~1781 m. Observations were made at SAAO because it is at high altitude in the desert and is free from air and light pollution, which allows the observer to record any TLE occurring within a radius of 900 km.

The observations were made during the high thunderstorm season in South Africa, which is from December to February every year. Although there was a lot of thunderstorm activity during the study, on most nights it was impossible to observe TLEs due to high local humidity, which makes it harder for a camera to detect TLE events, and local cloudy weather around SAAO, which obscures the camera's vision of distant clouds at a range of about 100–900 km. At times the thunderstorm was too close (<100 km) and at other times too far (>900 km) from the observation site to observe TLEs.[12]

The sprite videos were acquired using a Watec 910Hx camera together with a computer system.[12] The Watec 910Hx camera is a night-vision (sensitivity threshold of 0.64 milli-Lux), monochromatic, low-light level charge-coupled device television camera, which has been used worldwide to observe TLEs.[8,16,20-22] The 8.0 mm f/1.4 C-mount lens provided a field of view (FOV) of 35.3° horizontal and 26.6° vertical. The camera FOV and pointing direction were ascertained using background stars in the sprite video clips. The observation system captured the videos at 25 fps (frame period = 40 ms) with an image size of 352 × 288 pixels, and an angular resolution of 0.10/0.09°/pixel horizontal/vertical, and was operated with 8-bit intensity resolution.

The Watec 910Hx camera was connected to the computer via a video digitiser. The computer ran UFO-capture software version 2 (http://sonotaco.com/soft/e_index.html), which analyses the camera data in near real-time and detects all events with a detection threshold set by the operator. The UFO-capture software records a video clip for a few seconds from before until a few seconds after the triggering event. The UFO-capture software displays the time and date on the video screen. The computer was connected to the Network Timing Protocol server of SAAO, to ensure that the time was accurate (<1 ms). The camera viewing direction was approximately northeast of SAAO.[12]

The sprite observations reported here were observed from 18:43:12.0 until 21:38:06.5 UTC. Approximately 100 optical sprite events were observed during the study. However, of these, only 48 sprite events, for which the parent +CG lightning position information was available, were analysed to obtain altitude information (see Table 1). These ~100 sprites were classified as carrot (55%), carrot and column together (clustered sprites) (17%), column (13%), jellyfish (3%), unclassified (11%), and sprite halo (1%) as reported by Nnadih et al.[12] The 48 sprites analysed were categorised as carrot (37.5%), column (14.6%), clustered (25%), jellyfish (6.3%), unclassified sprites (14.6%) and sprite halo (2%). We did not observe any dancing sprites.

The lightning data were provided by the South African Weather Service (SAWS). The SAWS have 23 sensors located over the entire country. The SAWS sensors detect the cloud-to-ground lightning electromagnetic signals at very low frequency (1–10 kHz) and low frequency (100 kHz). The network systems use magnetic direction finding and time of arrival methods to find the angle between the sensor and lightning strike and to find the lightning geographical coordinates.[12,23]

Extremely low frequency signals recorded at the Schumann resonance station, located at Nagycenk Observatory (47.62° N,16.72° E), Central Europe, were used to confirm the polarity and estimate the CMC of the sprite parent lightning strokes. The Nagycenk Observatory is approximately 9000 km to the north from SAAO. The vertical electric and the two horizontal magnetic components of the atmospheric electromagnetic field in the ~3–30 Hz band are measured by a Schumann resonance monitoring system. Estimation of the CMC values used in this work is described in more detail by Nnadih et al.[12]

Observation technique and data analysis

The altitude of sprites was determined from their ground distance and elevation angle from the observation site. The ground distance to the projection of the sprites to the Earth's surface was assumed to be the distance to the parent lightning stroke. The elevation angle was obtained from the captured video by determining the exact pointing direction of the camera and then the direction of selected pixels of the sprite (see Supplementary appendix 1). A detailed description of the observation technique is given in Nnadih et al.[12]

Finding the direction of sprites

The azimuth and elevation angle of objects in the image and FOV of the camera can be obtained by recognising the stars in the background of a recorded image. This was achieved using the Stellarium software version 0.17 (https://stellarium.org/).

Knowing the stars' right ascension and declination from the star almanac, the time and camera location were used to calculate the stars' azimuth and elevation angles (see Supplementary appendix 1) within the image, plot the stars on the sprite image background and determine the camera FOV and pointing direction as well as the top altitude and maximum brightness altitude of the sprites (see Figure 1). The Python programming language was used to calculate the stars' azimuth and elevation angles from first principles (see Supplementary appendix 1) and plot the stars on the sprite image background. The calculated starfield can be manually adjusted within the image until a good fit is found to within one pixel. For the narrow FOV used, lens distortion did not need to be compensated for. However, atmospheric refraction was taken into account as described in Supplementary appendix 1.[24] The star fit allows the unique conversion of the image's pixel positions to azimuth and elevation angles. The video frame timestamp and the sprite image azimuth were used to identify the parent lightning from the SAWS lightning data. In addition, the large +CG lightning strikes required to initiate most sprites occur relatively rarely and are easily identified.

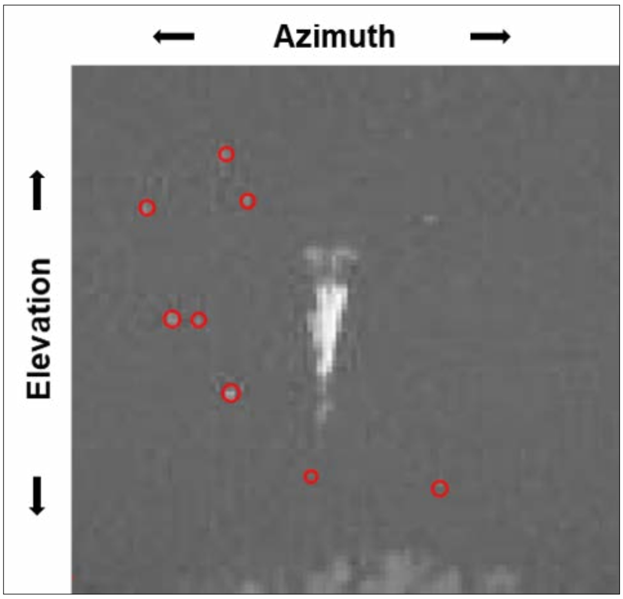


Figure 1: Sprite image with star fitting recorded on 11 January 2016 at 18:53:49.4 UTC at the South African Astronomical Observatory in Sutherland, South Africa. Red dots are the fitted stars.

Estimation of altitude

In order to estimate the altitude of a sprite, we applied spherical trigonometry in the horizontal plane and planar trigonometry in the vertical plane to determine the top altitude and the altitude of maximum brightness of sprites[25] (see Supplementary appendix 1). Equations 1 and 2 were used to determine the slant distance to a sprite (Equation 1) and the altitude of a sprite (Equation 2) (Figure 2).

$$r = \frac{2R_E\sin(\Theta) + \sqrt{4R_E^2\sin^2(\Theta) + 4\left[\frac{\cos^2(\Theta)-\sin^2(A)}{\sin^2(A)}\right]R_E^2}}{2\left[\frac{\cos^2(\Theta)-\sin^2(A)}{\sin^2(A)}\right]} \qquad \text{Equation 1}$$

$$h = \left[\frac{2R_E\sin(\Theta) + \sqrt{4R_E^2\sin^2(\Theta) + 4\left[\frac{\cos^2(\Theta)-\sin^2(A)}{\sin^2(A)}\right]R_E^2}}{2\left[\frac{\cos^2(\Theta)-\sin^2(A)}{\sin(A)}\right]}\right]\cos(\Theta) - R_E$$

Equation 2

where R_E is the radius of the Earth, r is the slant distance to a sprite, h is the altitude of a sprite, Θ is the elevation angle of a sprite from the camera position, and A is the great circle angular range.

In order to determine the altitude of the sprite from single-camera observations, we assumed that the sprite occurs directly above the parent lightning strike, whose location is known from the SAWS lightning data.[23] Most of the observed sprite elements were associated with a single lightning stroke, based on the time and position data. The observed sprite events without parent lightning information and those which were of unclear origin were not considered for the altitude analysis. We analysed the sprite videos frame by frame because it is possible to have more than one lightning strike and sprite event within one 40 ms video frame.[12]

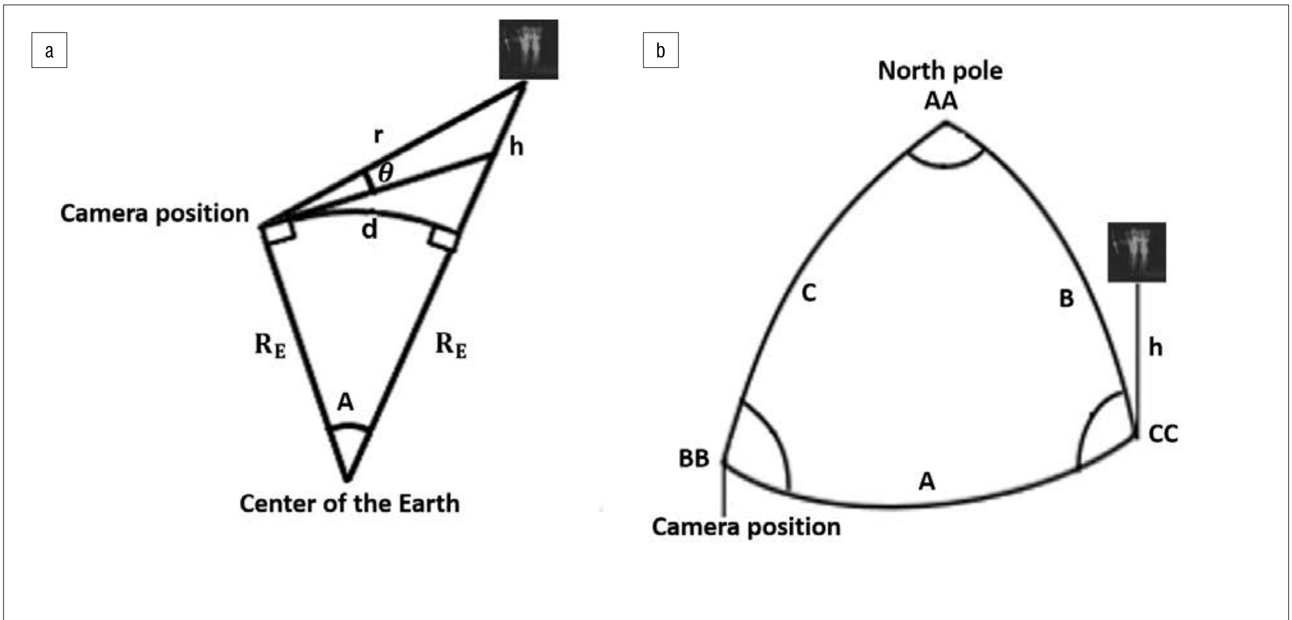
In cases in which the parent lightning initiated several sprite elements, we determined the top altitude and the altitude of maximum brightness of the first sprite element in the video frame. Where possible, the first sprite was identified by analysing the sprite initiation and subsequent development over several video frames. Sprite events for which we could not tell which sprite element initiated first were not included in the altitude analysis. In cases in which sprite elements generated by two lightning strokes appeared in one video frame, based on the lightning's latitude and longitude and the elevation and azimuth angle of the sprites in the video frame, we were able to determine which lightning strike initiated a certain sprite event. The azimuth angle of the lightning and sprites within the camera's FOV usually provide a clear indication of whether the lightning initiated the sprites or not, i.e. similar lightning and sprite azimuth angles in the same image frame imply association and vice versa.

Although there is evidence that sprites can be displaced horizontally from the parent lightning strike[14-16,26], our assumption that they are co-located is reasonable because there is evidence (optical observations) of sprites occurring over the parent lightning[15,16]. However, the most probable production mechanism for sprites suggests that in fact there is a horizontal displacement in most of the cases (typically up to 21 km).[27,28] A shift of tens of kilometres is small in comparison to the camera–sprite distance of ~440–890 km in our data set, so the displacement effect is small and would affect only the calculations along the radial line of sight from the point of observation.

Error analysis

The uncertainty in the estimation of altitude depends on the luminosity limit of the camera and the accuracy of the star fitting, which affects the sprite elevation angle estimation.[15,16] The uncertainty of the estimate of altitude also depends on the distance from the sprite to the observer. The angular resolution for the camera used during the study was 0.10/0.09°/pixel horizontal/vertical, respectively, which implies that one pixel is approximately 0.1°.[19] This value is comparable to that used by Stenbaek-Nielsen et al.[16] in their triangulation study of sprites.

In order to obtain the top altitude of a sprite and the altitude of maximum brightness of a sprite on a calibrated sprite image, we selected a point on the sprite to obtain the elevation angle and used Equation 2 to calculate the altitude. The pixel corresponding to the top altitude of the sprite is the pixel of greatest elevation angle within the sprite that has an intensity value greater than the background intensity for the same azimuth angle.



d = great circle range; R_E = radius of the Earth; r = slant distance to a sprite; h = altitude of a sprite; θ = elevation angle of a sprite from the camera position; B = co-latitude of sprite; C = co-latitude of the camera; AA = change in longitude between sprite and camera position; BB = azimuth angle of a sprite from the camera position; A = great circle range on the ground over the radius

Figure 2: Schematic representation of basic triangles on a sphere: (a) altitude of the sprite vertical axis and (b) horizontal axes.

The altitude of maximum brightness of a sprite was determined by identifying the brightest pixel within a sprite image.[29] This was done by sampling all the pixels within the sprite. The altitude error was obtained by calculating the altitude of two points 0.1° (one pixel) above and below the point used to determine the target altitude. The change in altitude was then determined with respect to the top altitude and the altitude of maximum brightness of the sprite. This change in altitude depends on the slant distance from the top altitude or altitude of maximum brightness of the sprite to the camera (r).

The sprite events we observed did not saturate the camera's detector because of the settings used, therefore we were able to determine the elevation angle corresponding to the top altitude and altitude of maximum brightness sprite pixel to an accuracy of one pixel. The time delay between the parent lightning and the sprites cannot be measured to better than 40 ms, i.e. one video frame, the best optical time resolution available. Therefore, for our temporal resolution, we could not determine a time delay between the sprite and the parent lightning strike.

Results

The 48 sprites analysed consisted of 18 carrot, 7 column, 12 clustered (carrot and column), 3 jellyfish and 7 unclassified sprites as well as 1 sprite halo. From the SAWS lightning data, all observed sprite events were produced by +CG lightning strikes with a peak current magnitude varying from 11 kA to 191 kA.[12]

The star fitting revealed that the sprite image axes azimuth and elevation angles ranged over 30.5–65.8° and 0.1–26.7°, respectively. The slant distance from the observer's location to the top altitude of the 48 sprite event locations was found to be 439.93–898.02 km, giving an altitude error estimate of ±0.35–0.8 km, respectively, for one-pixel accuracy. The slant distance from the observer's location to the altitude of maximum brightness of the 48 sprite event locations was found to be 437.02–896.0 km, giving an altitude error estimate of ±0.29–0.74 km, respectively (see Table 1), again for one-pixel accuracy. Repeating the analyses for two-pixel accuracy increases the altitude estimate by about 0.8 km and the altitude uncertainty to ±0.8–1.5 km. These findings are consistent with those of Stenbaek-Nielsen et al.[16]

Distribution of top sprite altitudes

Figure 3 shows the 48 sprites' top altitude determined through Equation 2 with respect to the time at which they occurred. The different sprite morphologies are denoted by different colours in Figure 3.

The top altitude of the sprites indicates the altitude at which the strength of the electric field in the given sprite is above a particular level, which is generally lower than the critical breakdown electric field at that altitude.[1,30] From our analysis, we found that the top altitude of sprites ranged from 73.6 km to 95.6 km above the Earth's surface. The sprites' top altitude is not static; however, it does not depend on the time at which they occurred. The average top altitude of sprites was found to be approximately 84.3 km with a standard deviation of 5.5 km. The average top altitudes for column, carrot, clustered, jellyfish and unclassified sprites were found to be about 83.2, 84, 86, 90.7 and 82 km, with the standard deviation of 6.0, 5.5, 4.2, 2.1 and 6.4 km, respectively. Our results are consistent with the reports of Sentman et al.[14] and Stenbaek-Nielsen et al.[16] and the top altitude of column and carrot sprites reported by Wescott et al.[15] and Bor[31].

Distribution of the altitudes of maximum sprite brightness

The altitudes of maximum brightness for the 48 sprites are shown in Figure 4. The altitude of the maximum brightness of sprites is that at which the photon production maximises. This region has low conductivity, and the electrons are converted to negative ions which intensifies the electric field and maximises photon production.[32] From our analysis, we found that the altitude of maximum brightness ranges between 53.5 km and 84.1 km. The average maximum brightness altitude was found to be about 69 km with a standard deviation of 6.2 km. These results are consistent with those of Luque et al.[30] The average altitudes of maximum brightness for column, carrot, clustered, jellyfish and unclassified sprites were found to be about 67.4, 67.3, 70, 74.3 and 73.3 km, with corresponding standard deviations of 7.6, 6.0, 7.7, 1.2 and 4.4 km, respectively. The altitudes of maximum brightness for column and carrot sprites are in agreement with those reported in the literature.[15,31,32]

The altitude of maximum brightness also does not depend on the time of the event. From the literature, the altitude of maximum brightness region lies between 65 km and 85 km.[1,14,33] However, we found that some sprites have a maximum brightness altitude lower than 65 km above the Earth's surface, with the lowest observed being at 53.5 km, which is consistent with Füllekrug et al.'s[34] findings. These results therefore suggest that the photon production maximises at an altitude range of about 53.5–84.1 km, with an average of 69 km.

Sprite altitude versus charge moment change

From Schumann resonance observations[4,35,36] we were able to compute the lightning CMC values associated with 9 of the 48 sprites (Table 1). Figure 5 shows the lightning CMC values associated with these nine sprite events plotted against the top altitude. The +CG lightning strikes, which initiated sprites with a top altitude ranging from 83.7 km to 92.0 km, yielded a CMC ranging from 900 C km to 2100 C km. There is a positive correlation (correlation coefficient of 0.6) between the estimated top altitude of sprites and lightning CMC.[37] For the altitude of maximum brightness versus CMC (not shown), there is a moderate positive correlation (correlation coefficient of 0.38). This data set is small and therefore more data are needed to reliably confirm the inferred trend.

Conclusions

Our analysis of the top altitudes and altitudes of maximum brightness of sprites recorded for the first time in South Africa shows good agreement with previous sprite observations reported in the literature.

The average estimated top altitude for the 48 sprites analysed is approximately 84.3 km with a standard deviation of about 5.5 km. The top altitude of those 48 sprites ranged from 73.6 km to 95.6 km. The top altitude uncertainty was found to vary from ±0.35 km to 0.8 km, depending on the slant distance from the observer site to the top of the sprite, which was found to be between approximately 440 km and 898 km.

The maximum brightness region of sprites was found to be in the altitude range of about 53.5–84.1 km. The average altitude of maximum brightness was approximately 69 km, with a standard deviation of 6.2 km. The slant distance to the maximum brightness part of the sprite from the observer site ranged from approximately 437 km to 896 km and the altitude uncertainty was found to vary from ±0.29 km to 0.74 km.

The linear correlation between the estimated top altitude of sprites and lightning CMC is equal to 0.6 and may be regarded as a moderately high correlation, although the data set is limited. A moderate correlation of 0.38 exists between the altitude of maximum brightness of sprites and the lightning CMC.

Acknowledgements

D.C.M. is grateful to the South African National Space Agency and the University of the Western Cape for the support they provided. We thank the South African Weather Service for their assistance in providing the cloud-to-ground lightning activity data. We also thank the South African Astronomical Observatory for allowing the sprites campaign team to conduct the study at their facility in Sutherland, Northern Cape Province, South Africa. The contribution of J.B. was supported by the National Research, Development and Innovation Office, Hungary-NKFIH, K115836. J.B. also appreciates the support of COST Action CA15211, ELECTRONET, in facilitating scientific communication.

Competing interests

We declare that there are no competing interests.

Table 1: Sprite initiation time (coordinated universal time; UTC), lightning location, lightning peak current, distances, sprites' top altitude and altitude of maximum brightness. Lightning data were provided by the South African Weather Service. The charge moment change (CMC) data were computed from the Schumann resonance station at the Nagycenk Observatory, Hungary (47.62° N, 16.72° E). The sprites' top altitudes and altitudes of maximum brightness were estimated as described in the text.

Sprite initiation time (UTC) (±20 ms)	Longitude (°)	Latitude (°)	Peak current (kA)	Distance from the camera to the lightning location (km)	Slant distance from the camera to the top altitude (km)	Slant distance from the camera to the maximum brightness (km)	Top altitude of sprite (km)	Altitude of maximum brightness of sprite (km)	CMC (C km)
18:53:49.4	25.34	-31.82	37	432.18	439.93	437.02	80.2	68.8	
18:59:10.5	25.55	-31.90	82	450.25	458.52	456.94	76.3	68.2	
19:10:04.6	24.71	-27.94	58	619.94	629.36	628.17	79.2	76.3	
19:12:54.9	24.32	-28.51	29	546.42	556.63	553.88	82.3	67.1	
19:17:09.8	24.63	-28.31	23	582.53	613.27	612.43	76.4	73.1	
19:17:09.8	24.42	-28.17	46	582.56	591.54	590.93	76.6	72	
19:17:09.8	24.76	-28.15	23	604.36	592.27	588.39	88.9	65	
19:22:25.3	24.02	-28.52	97	528.24	537.97	533.24	83.1	55.1	1600
19:23:10.2	23.38	-28.12	51	534.03	545.29	541.29	89.5	70	
19:28:34.2	24.65	-27.62	35	645.51	654.88	651.84	91.3	83.5	
19:28:34.2	24.65	-27.62	55	645.51	654.84	654.34	91.5	74.4	
19:28:34.2	23.44	-28.30	69	519.11	582.98	581.36	80.5	72.9	
19:33:24.0	24.79	-27.88	189	629.97	639.88	637.94	81.5	70.2	
19:34:46.6	23.79	-27.44	55	619.51	629.27	627.71	88	79	
19:36:07.3	24.24	-28.23	72	566.94	577.77	575.56	83.2	71.6	
19:38:32.5	23.55	-27.54	48	599.09	609.11	606.11	87.2	72.9	1000
19:48:05.3	25.14	-28.22	47	622.09	629.84	627.4	79.2	76.5	
19:48:05.3	25.03	-27.70	93	660.1	669.78	667.5	90.1	74.5	
19:49:19.7	24.09	-28.00	56	580.08	589.07	587.3	79.6	74.9	1500
19:49:54.3	27.04	-27.48	57	810.87	818	815	84.8	65	
19:52:23.9	24.08	-28.26	87	555.36	566.49	561.72	86.9	66.5	
19:52:35.4	26.98	-27.62	74	795.91	834.33	830.56	85.6	62.7	
19:55:08.1	24.49	-27.69	17	630.3	639.6	637.84	78.9	68.5	
19:55:08.1	24.46	-27.54	91	642.66	652.67	651.2	82.6	74.3	
19:58:01.3	27.05	-27.68	42	796.46	804.83	802.27	79.2	62	
19:58:29.6	25.12	-27.69	11	666.36	677.84	676.58	90.8	84.1	1800
20:00:32.6	26.72	-27.31	37	801.72	811.45	807.02	82.8	53.5	
20:08:03.3	23.94	-28.24	88	549.73	559.35	557.9	77.4	69.5	
20:09:25.3	23.63	-27.91	71	566.02	577.54	574.24	87.9	70.2	1200
20:10:04.9	24.66	-28.18	90	595.72	605.56	601.92	82.1	60.7	
20:10:04.9	23.96	-28.26	101	548.91	558.89	555.68	80.6	62.4	
20:14:02.3	24.63	-27.28	105	675.99	684.63	683.52	73.6	66.5	
20:16:22.7	27.08	-27.67	26	799.39	807.52	805.13	77	60.8	
20:18:46.4	24.84	-28.06	104	617.07	628.7	624.72	92.2	69.9	2000
20:22:27.6	23.94	-28.26	33	547.86	558.64	555.89	84.9	69.9	
20:24:43.7	24.64	-27.56	191	650.48	661.22	658.63	86.4	71.2	
20:26:21.9	27.57	-27.55	117	844.34	854.19	851.76	88.3	73.1	
20:27:34.6	24.64	-27.43	20	662.52	674.98	670.55	95.6	70.2	
20:29:53.8	24.66	-27.51	126	656.19	668.04	663.56	92.3	75.5	2000
20:33:19.2	23.95	-27.56	101	615.14	626.6	621.82	88.3	59.7	
20:38:43.4	27.43	-27.57	85	832.49	841.36	840.28	82.1	75.2	
20:46:42.0	27.78	-27.92	80	834.15	834	832.52	76.3	62.2	
20:46:42.0	27.73	-27.97	87	826.86	843.2	838.68	84.8	62.7	
21:01:08.0	23.91	-26.40	48	728.84	741.02	736.37	91.7	62.6	
21:29:05.4	25.01	-26.73	46	747.66	757.78	755.79	83.7	71.2	900
21:33:48.1	25.09	-26.74	159	750.97	761.15	759.38	83.1	72.1	
21:38:06.0	27.42	-26.83	31	888.26	898.02	896	91.6	72.7	
21:41:24.3	27.70	-27.44	117	862.04	872.53	869.85	92	75.6	2100
Average altitude		84.3			69				
Standard deviation		5.5			6.2				

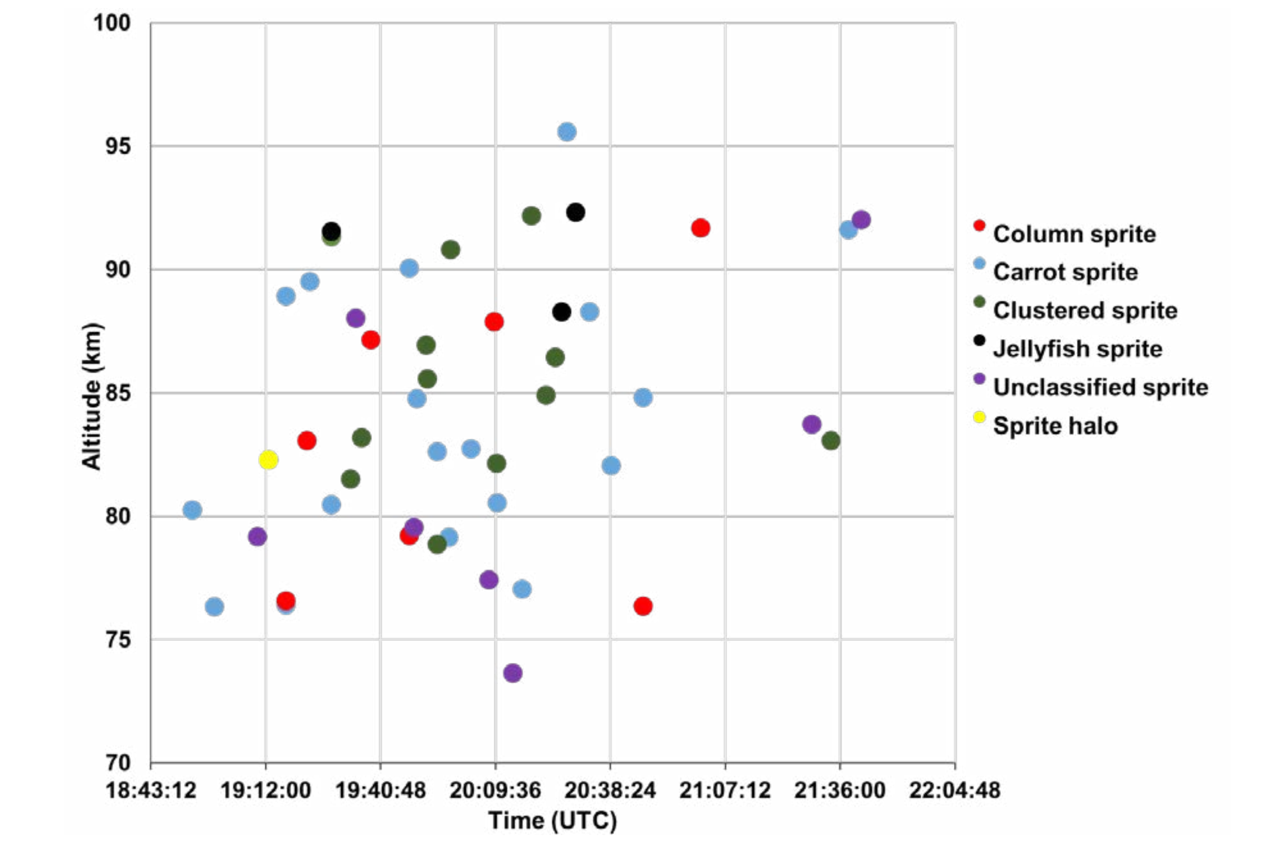


Figure 3: The top altitude of the 48 sprites with respect to the time at which they were observed during the 2016 sprites campaign.

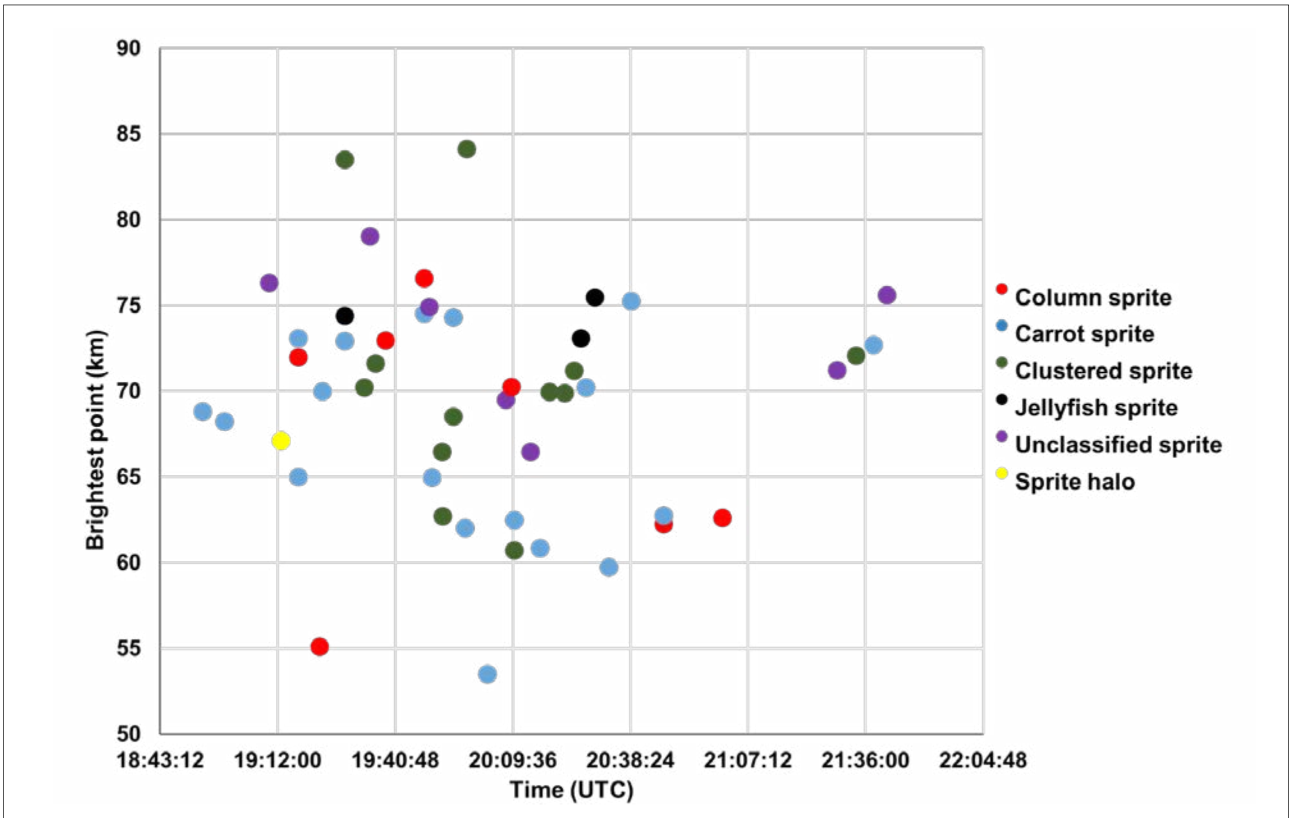


Figure 4: The altitude of maximum brightness of the 48 sprites with respect to their time of occurrence.

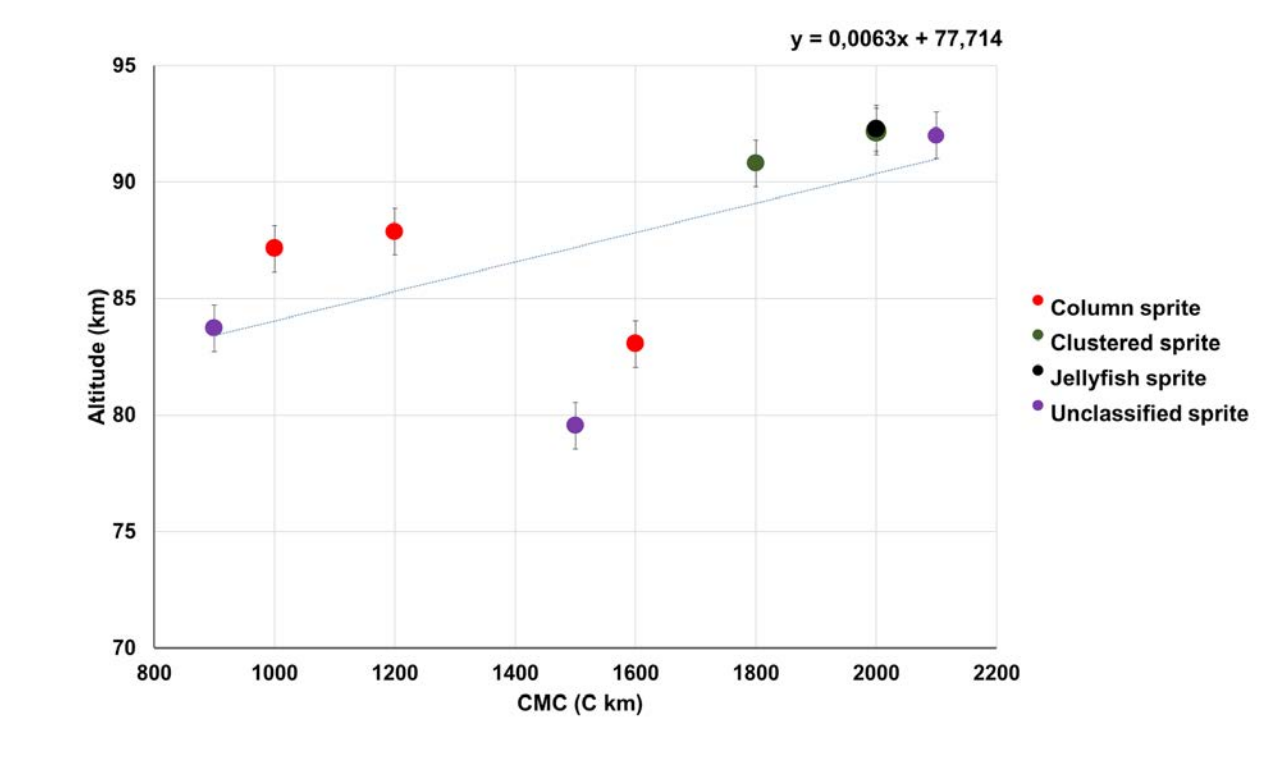


Figure 5: The relationship between charge moment change (CMC) and the top altitude of sprites.

Authors' contributions

D.C.M.: Conceptualisation, methodology, data collection, sample analysis, data analysis, validation, data curation, writing – the initial draft, writing – revisions. M.J.K.: Conceptualisation, methodology, data collection, sample analysis, validation, student supervision, project leadership, project management. J.B.: Data collection, sample analysis. S.N.: Data collection, sample analysis.

References

1. Siingh D, Singh R, Singh A, Kumar S, Kulkarni M, Singh A. Discharges in the Stratosphere and Mesosphere. Space Sci Rev. 2012;169(1–4):73–121. https://doi.org/10.1007/s11214-012-9906-0
2. Füllekrug M, Rycroft M. The contribution of sprites to the global atmospheric electric circuit. Earth Planets Space. 2006;58(9):1193–1196. https://doi.org/10.1186/bf03352009
3. Rycroft M, Odzimek A. Effects of lightning and sprites on the ionospheric potential, and threshold effects on sprite initiation, obtained using an analog model of the global atmospheric electric circuit. J Geophys Res. 2010;115(A6), A00E37. https://doi.org/10.1029/2009ja014758
4. Soula S, Iacovella F, Van der Velde O, Montanyà J, Füllekrug M, Farges T, et al. Multi-instrumental analysis of large sprite events and their producing storm in southern France. Atmos Res. 2014;135–136:415–431. https://doi.org/10.1016/j.atmosres.2012.10.004
5. Soula S, Mlynarczyk J, Füllekrug M, Pineda N, Georgis J, Van der Velde O, et al. Dancing sprites: Detailed analysis of two case studies. J Geophys Res Atmos. 2017;122(6):3173–3192. https://doi.org/10.1002/2016jd025548
6. Pasko V, Yair Y, Kuo C. Lightning related transient luminous events at high altitude in the Earth's atmosphere: Phenomenology, mechanisms and effects. Space Sci Rev. 2011;168(1–4):475–516. https://doi.org/10.1007/s11214-011-9813-9
7. Lyons W. Sprites. Encyclopedia of atmospheric sciences. Atmos Res. 2003;3:738–747. https://doi.org/10.1016/b0-12-227090-8/00146-9
8. Yaniv R, Yair Y, Price C, Bór J, Sato M, Hobara Y, et al. Ground-based observations of the relations between lightning charge-moment-change and the physical and optical properties of column sprites. J Atmos Sol Terr Phys. 2014;107:60–67. https://doi.org/10.1016/j.jastp.2013.10.018
9. Neubert T, Allin T, Blanc E, Farges T, Haldoupis C, Mika A, et al. Co-ordinated observations of transient luminous events during the EuroSprite2003 campaign. J Atmos Sol Terr Phys. 2005;67(8–9):807–820. https://doi.org/10.1016/j.jastp.2005.02.004
10. Franz R, Nemzek R, Winckler J. Television Image of a large upward electrical discharge above a thunderstorm system. Science. 1990;249(4964):48–51. https://doi.org/10.1126/science.249.4964.48
11. Hu W, Cummer S, Lyons W, Nelson T. Lightning charge moment changes for the initiation of sprites. Geophys Res Lett. 2002;29(8):120-1–120-4. https://doi.org/10.1029/2001gl014593
12. Nnadih S, Kosch M, Martinez P, Bor J. First ground-based observations of sprites over southern Africa. S Afr J Sci. 2018;114(9/10), Art. #4272. https://doi.org/10.17159/sajs.2018/4272
13. Chen A, Kuo C, Lee Y, Su H, Hsu R, Chern J, et al. Global distributions and occurrence rates of transient luminous events. J Geophys Res. 2008;113(A8), A08306. https://doi.org/10.1029/2008ja013101
14. Sentman D, Wescott E, Osborne D, Hampton D, Heavner M. Preliminary results from the Sprites94 Aircraft campaign: 1. Red sprites. Geophys Res Lett. 1995;22(10):1205–1208. https://doi.org/10.1029/95gl00583
15. Wescott E, Sentman D, Heavner M, Hampton D, Lyons W, Nelson T. Observations of 'Columniform' sprites. J Atmos Sol Terr Phys. 1998;60(7–9):733–740. https://doi.org/10.1016/s1364-6826(98)00029-7
16. Stenbaek-Nielsen H, Haaland R, McHarg M, Hensley B, Kanmae T. Sprite initiation altitude measured by triangulation. J Geophys Res. 2010;115(A3), A00E12. https://doi.org/10.1029/2009ja014543
17. Ganot M, Yair Y, Price C, Ziv B, Sherez Y, Greenberg E, et al. First detection of transient luminous events associated with winter thunderstorms in the eastern Mediterranean. Geophys Res Lett. 2007;34(12), L12801, 5 pages. https://doi.org/10.1029/2007gl029258

18. Li J, Cummer S, Lyons W, Nelson T. Coordinated analysis of delayed sprites with high-speed images and remote electromagnetic fields. J Geophys Res. 2008;113(D20), D20206, 11 pages. https://doi.org/10.1029/2008jd010008

19. McHarg M, Stenbaek-Nielsen H, Kammae T. Observations of streamer formation in sprites. Geophys Res Lett. 2007;34(6), L06804, 5 pages. https://doi.org/10.1029/2006gl027854

20. Soula S, Van der Velde O, Montanyà J, Neubert T, Chanrion O, Ganot M. Analysis of thunderstorm and lightning activity associated with sprites observed during the EuroSprite campaigns: Two case studies. Atmos Res. 2009;91(2–4):514–528. https://doi.org/10.1016/j.atmosres.2008.06.017

21. Gamerota W, Cummer S, Li J, Stenbaek-Nielsen H, Haaland R, McHarg M. Comparison of sprite initiation altitudes between observations and models. J Geophys Res. 2011;116(A2), A02317, 9 pages. https://doi.org/10.1029/2010ja016095

22. Liu N, Boggs L, Cummer S. Observation-constrained modeling of the ionospheric impact of negative sprites. Geophys Res Lett. 2016;43(6):2365–2373. https://doi.org/10.1002/2016gl068256

23. Gijben M. The lightning climatology of South Africa. S Afr J Sci. 2012;108(3/4), Art. #740, 10 pages. https://doi.org/10.4102/sajs.v108i3/4.740

24. Duffett-Smith P. Practical astronomy with your calculator. 3rd ed. Cambridge: Cambridge University Press; 1989. https://doi.org/10.1017/CBO9780511564895

25. Siranah.de. Applications of spherical trigonometry in navigation [webpage on the Internet]. No date [updated 2010 Dec 03; cited 2019 Nov 20]. Available from: http://www.siranah.de/html/sail042e.htm

26. Lyons W. Sprite observations above the U.S. high plains in relation to their parent thunderstorm systems. J Geophys Res Atmos. 1996;101(D23):29641–29652. https://doi.org/10.1029/96jd01866

27. Mlynarczyk J, Bór J, Kulak A, Popek M, Kubisz J. An unusual sequence of sprites followed by a secondary TLE: An analysis of ELF radio measurements and optical observations. J Geophys Res Space Phys. 2015;120(3):2241–2254. https://doi.org/10.1002/2014ja020780

28. Bór J, Zelkó Z, Hegedüs T, Jäger Z, Mlynarczyk J, Popek M, et al. On the series of +CG lightning strokes in dancing sprite events. J Geophys Res Atmos. 2018;123(19):11030–11047. https://doi.org/10.1029/2017jd028251

29. Pasko V, Inan U, Bell T. Spatial structure of sprites. Geophys Res Lett. 1998;25(12):2123–2126. https://doi.org/10.1029/98gl01242

30. Luque A, Stenbaek-Nielsen H, McHarg M, Haaland R. Sprite beads and glows arising from the attachment instability in streamer channels. J Geophys Res Space Phys. 2016;121(3):2431–2449. https://doi.org/10.1002/2015ja022234

31. Bór J. Optically perceptible characteristics of sprites observed in central Europe in 2007–2009. J Atmos Sol Terr Phys. 2013;92:151–177. https://doi.org/10.1016/j.jastp.2012.10.008

32. Malagón-Romero A, Teunissen J, Stenbaek-Nielsen H, McHarg M, Ebert U, Luque A. On the emergence mechanism of carrot sprites. Geophys Res Lett. 2020;47(1), e2019GL085776, 8 pages. https://doi.org/10.1029/2019gl085776

33. Neubert T, Rycroft M, Farges T, Blanc E, Chanrion O, Arnone E, et al. Recent results from studies of electric discharges in the Mesosphere. Surv Geophys. 2008;29(2):71–137. https://doi.org/10.1007/s10712-008-9043-1

34. Füllekrug M, Nnadih S, Soula S, Mlynarczyk J, Stock M, Lapierre J, et al. Maximum sprite streamer luminosity near the Stratopause. Geophys Res Lett. 2019;46(21):12572–12579. https://doi.org/10.1029/2019gl084331

35. Williams E, Lyons W, Hobara Y, Mushtak V, Asencio N, Boldi R, et al. Ground-based detection of sprites and their parent lightning flashes over Africa during the 2006 AMMA campaign. Quart J Roy Meteor Soc. 2009;136(S1):257–271. https://doi.org/10.1002/qj.489

36. Ádám A, Bencze P, Bór J, Heilig B, Kis Á, Koppán A, et al. Geoelectromagnetism and the changing Earth. Acta Geod Geophys Hung. 2009;44(3):271–312. https://doi.org/10.1556/ageod.44.2009.3.3

37. Peat J, Barton B, Elliott E. Statistics workbook for evidence-based health care. New York: John Wiley & Sons; 2009.

The North-West University's High Altitude Radiation Monitor programme

AUTHORS:
Moshe G. Mosotho[1,2]
Roelf D. Strauss[1]
Rendani R. Nndanganeni[2]
Jacobus P. van den Berg[1,2]

AFFILIATIONS:
[1]Centre for Space Research, North-West University, Potchefstroom, South Africa
[2]Space Science Division, South African National Space Agency, Hermanus, South Africa

CORRESPONDENCE TO:
Moshe Mosotho

EMAIL:
mosothomoshegodfrey@gmail.com

DATES:
Received: 30 Oct. 2019
Revised: 25 Feb. 2020
Accepted: 13 June 2020
Published: 29 Jan. 2021

HOW TO CITE:
Mosotho MG, Strauss RD, Nndanganeni RR, van den Berg JP. The North-West University's High Altitude Radiation Monitor programme. S Afr J Sci. 2021;117(1/2), Art. #7561. https://doi.org/10.17159/sajs.2021/7561

ARTICLE INCLUDES:
☒ Peer review
☐ Supplementary material

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

EDITOR:
Amanda Weltman

KEYWORDS:
radiation, atmosphere, dose rate, aircraft, public awareness

FUNDING:
South African National Space Agency, South African National Research Foundation, South African National Astrophysics and Space Science Programme

Since the discovery of cosmic radiation by Victor Hess in 1912, when he reported a significant increase in radiation as altitude increases, concerns about radiation effects on human bodies and equipment have grown over the years. The secondary and tertiary particles which result from the interaction of primary cosmic rays with atmospheric particles and commercial aircraft components, are the primary cause of the radiation dose deposited in human bodies and in electronic equipment (avionics) during aircraft flights. At an altitude of about 10 km (or higher) above sea level, the dose received by frequent flyers, and especially flight crew, is a serious concern. Also of concern is the possible failure of sensitive equipment on board commercial aircrafts as a result of flying through this mixed radiation field. Monitoring radiation in the atmosphere is therefore very important. Here we report on the first measurements by the High Altitude Radiation Monitor (HARM) detector during a commercial flight from Johannesburg (O.R. Tambo International Airport) to Windhoek (Hosea Kutako International Airport). As part of a public awareness activity, the HARM detector was placed on a high-altitude balloon, and these measurements are also shown here. Model calculations (estimations) of radiation levels for the commercial aircraft flight are shown and the results are used to interpret our measurements.

Significance:

- Measurements of the Regener–Pfotzer maximum in South Africa and dosimetric measurements on board a commercial flight are presented.
- These radiation measurements are compared to model calculations which can be used to predict the radiation dose during commercial flights.
- This study also aims to raise public awareness about the atmospheric radiation environment from ground level to the Regener–Pfotzer peak at high altitude.

Introduction

Earth is continuously exposed to ionising radiation from high-energy particles originating from galactic sources such as active galactic nuclei or supernova remnants[1-3], the Sun[4] and particles trapped in the Van Allen radiation belts[5,6]. Galactic and solar particles are energetic enough to interact with the Earth's atmosphere where they produce high-energy secondary particles. During commercial aircraft flights, these primary and secondary high-energy particles interact with aircraft components to further produce more high-energy secondary particles. Scientists have been assessing the exposure of crew members and commercial aircraft passengers to ionising radiation ever since it was identified as a health risk in the 1920s[7], and since then, progress has been made over the years in measuring and monitoring radiation at flight altitudes[8,9]. When exposed to radiation from these high-energy particles, the time-dependent amount of energy absorbed by human bodies, referred to as the dose rate, can be harmful because it can damage the body's DNA make-up.[10,11] Even though our bodies are capable of repairing this damage, errors can occur during the repair processes, leading to a mutation of cells that can potentially result in cancer.[12,13] Therefore, measurements of these secondary particles can be of help to improve our knowledge and understanding of post-exposure effects and thereby also improve the risk estimation for commercial aircrafts.

The production of secondary particles in the upper atmosphere is at a maximum rate[14] at an altitude reported to be between 14 km and 20 km above sea level. This maximum altitude of ionising radiation is known as the Regener–Pfotzer maximum (RP-max).[15] The RP-max varies with geomagnetic (omnidirectional) cut-off rigidity, air pressure, and solar cycle activity and there is no fixed altitude position above sea level to represent it. Below this altitude, the intensity of secondary radiation steadily decreases due to absorption and decay processes.[16,17] Therefore, to measure and monitor radiation exposure, at least up to the RP-max, a small, portable and lightweight instrument known as the High Altitude Radiation Monitor (HARM) was used. This acronym was used to honour the late Prof. Harm Moraal, who was part of the North-West University's (NWU) physics department for 44 years and championed cosmic ray research in South Africa. This instrument uses a Geiger–Müller (GM) counter, capable of detecting secondary and tertiary particles, and a number of smaller sensors added to provide additional information.

The South African National Space Agency's (SANSA) Space Science Division operates a wide range of infrastructure, all dedicated to studying the atmosphere, the Earth's magnetic field, the Sun, and the near-space environment. The agency's ground-based instruments are located across southern Africa and in Antarctica. For years, SANSA has been conducting real-time monitoring and forecasting of space weather events (http://spaceweather.sansa.org.za). Recently, the agency was selected by the International Civil Aviation Organization (ICAO) as one of two regional centres to provide space weather services, including solar storm forecasts and warnings, to the aviation industry from 2022. By the end of 2018, the ICAO had recommended that all flight plans must include space weather information by law.[18] In anticipation of this, the HARM programme was initiated to characterise the radiation environment above South Africa at aviation altitudes. Here we report on the first HARM measurements (and the model estimates) during a commercial aircraft flight and the measurements obtained during a high-altitude balloon launch.

Cosmic rays in the atmosphere

Cosmic rays are energetic charged particles, consisting mostly of protons, which can be classified into four main species, namely solar energetic particles (SEPs), anomalous cosmic rays, Jovian electrons (which are produced in the Jovian magnetosphere), and galactic cosmic rays (GCRs). However, only the GCR and SEP species are of importance here. The SEPs originate from the Sun and are accelerated mainly by solar flares, coronal mass ejections, and shocks in the interplanetary medium. The source of the GCRs is, however, located outside the region of space influenced by the solar wind and the solar magnetic field, known as the heliosphere[19], in astrophysical sources such as supernova remnants[1-3]. As they enter the heliosphere with energies of up to 10^{12} GeV[20], these particles interact with the solar wind and the embedded heliospheric magnetic field, causing their intensities to change as a function of time, position, and energy in a process called cosmic ray modulation.[20] This process is classified based on the timescale in which it occurs, i.e. short-term (from hours to days[4]), medium-term (from days to about a month and sometimes more than that[21]), and long-term (from a month to years or decades and even beyond) modulation effects.

The level of modulation greatly depends on the energy of these particles. At Earth, cosmic rays are denied free access to the ground level due to the orientation of the geomagnetic field. Therefore, depending on their energies (or rigidity, which defines a measure of the momentum of the particle per elementary charge), their arrival direction, and the period of solar activity, these particles can be deflected back into space, get trapped in the geomagnetic field, or they can penetrate and make their way to the top of the atmosphere.

For vertically incident particles originating from space, it is easy for particles with both high and low energies to penetrate the geomagnetic field at the polar regions because the geomagnetic field lines are directed perpendicular to the surface, leading to higher fluxes in these regions. In a geocentric dipole field, geomagnetic cut-off rigidity is defined as the minimum momentum per charge a particle must have in order to reach a specific position, given a certain arrival direction.[22] Throughout this article, we use the geomagnetic cut-off rigidity of a particle with vertical incidence, labelled Pc. At Earth, Pc varies from 0 GV at the geomagnetic polar regions to about 17 GV in the mid-latitudes. Due to the more effective shielding at the mid-latitude regions, fewer particles can reach the ground level at mid-latitudes.[22,23]

As primary GCRs and SEPs penetrate the geomagnetic field and traverse down to lower altitudes they will interact with the atmospheric particles, producing secondary particles such as neutrons, protons and electrons. A simplified schematic (not drawn to scale) illustrating the primary cosmic ray interaction with the upper atmosphere which results in a cascade of secondary particles is shown in Figure 1. This cascade continuously occurs until the particles lose all their energy and are finally absorbed in the lower atmosphere.[23] For a cosmic ray particle to sustain an atmosphere cascade which is observed at ground level, the primary particles need at least 1 GV in rigidity.

To study the behaviour of secondary cosmic rays on both long and short timescales, neutron monitors (NMs) are the most preferred detectors. NMs have been observing cosmic ray activity since 1951 when J.A. Simpson built the first monitor.[24] They measure the secondary particles' by-products, mainly neutrons and protons.[25] The four chosen NMs considered here are located in Antarctica (at the South African research base, SANAE), South Africa (Hermanus and Potchefstroom) and Namibia (Tsumeb), at cut-off rigidities (altitudes) ranging from about 0.86 GV (0.052 km) in Antarctica to about 9.15 GV (1.240 km) in Namibia. They are all operated by the Centre for Space Research of the NWU. The Centre has been involved with these types of observation in the southern hemisphere for over 60 years since the late 1950s.

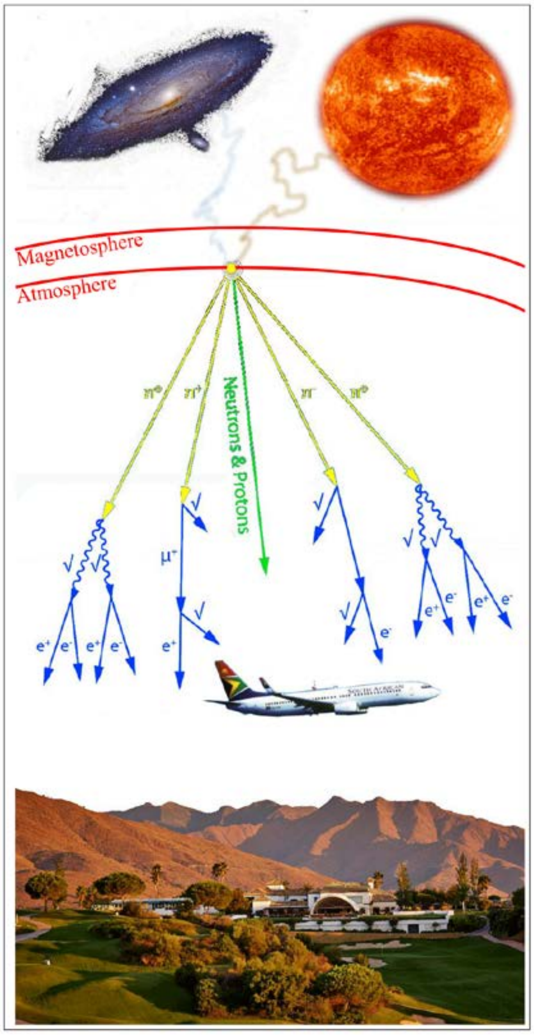


Figure 1: Basic representation of the cosmic ray cascade. The figure is not to scale.

Figure 2a shows the long-term cosmic ray NM count rates as a function of time, with all the NMs normalised to 100% in March 1987. Clearly, from this long-term record, NMs react to modulation based on Pc. A monitor with a lower Pc, i.e. SANAE at 0.86 GV, will be more sensitive to modulation and record higher count rates than a monitor with higher Pc, i.e. Tsumeb at 9.15 GV. Apart from detecting the secondary particles' by-products, NM counts on the ground can also give an indication of the primary cosmic ray flux at and above the top of the atmosphere.

In contrast to the quasi-constant quiet-time GCR levels, the Sun can produce short-term transient increases of cosmic rays at the ground level. These ground-level enhancements (GLEs) are caused by relativistic SEPs[26] being accelerated in solar flares and coronal mass ejections. These SEPs have sufficient intensity and energy to greatly enhance the atmospheric secondary radiation environment.[27] The first GLE was registered on 28 February 1942, and only 72 GLEs have occurred to date. Note that these solar transient events can also modulate the GCR levels over longer timescales.

Figure 2b shows the percentage increase measured by the four chosen NMs during the GLE of 29 September 1989, also referred to as GLE 42. This GLE is the only one for which all four NMs simultaneously registered a significant increase. The SANAE NM registered a 297.5% peak increase, while the Hermanus, Potchefstroom and Tsumeb NMs recorded 105.1%, 105.3% and 62.2% increases, respectively. A GLE can therefore lead to increases in the NM count rate of several orders of magnitude.

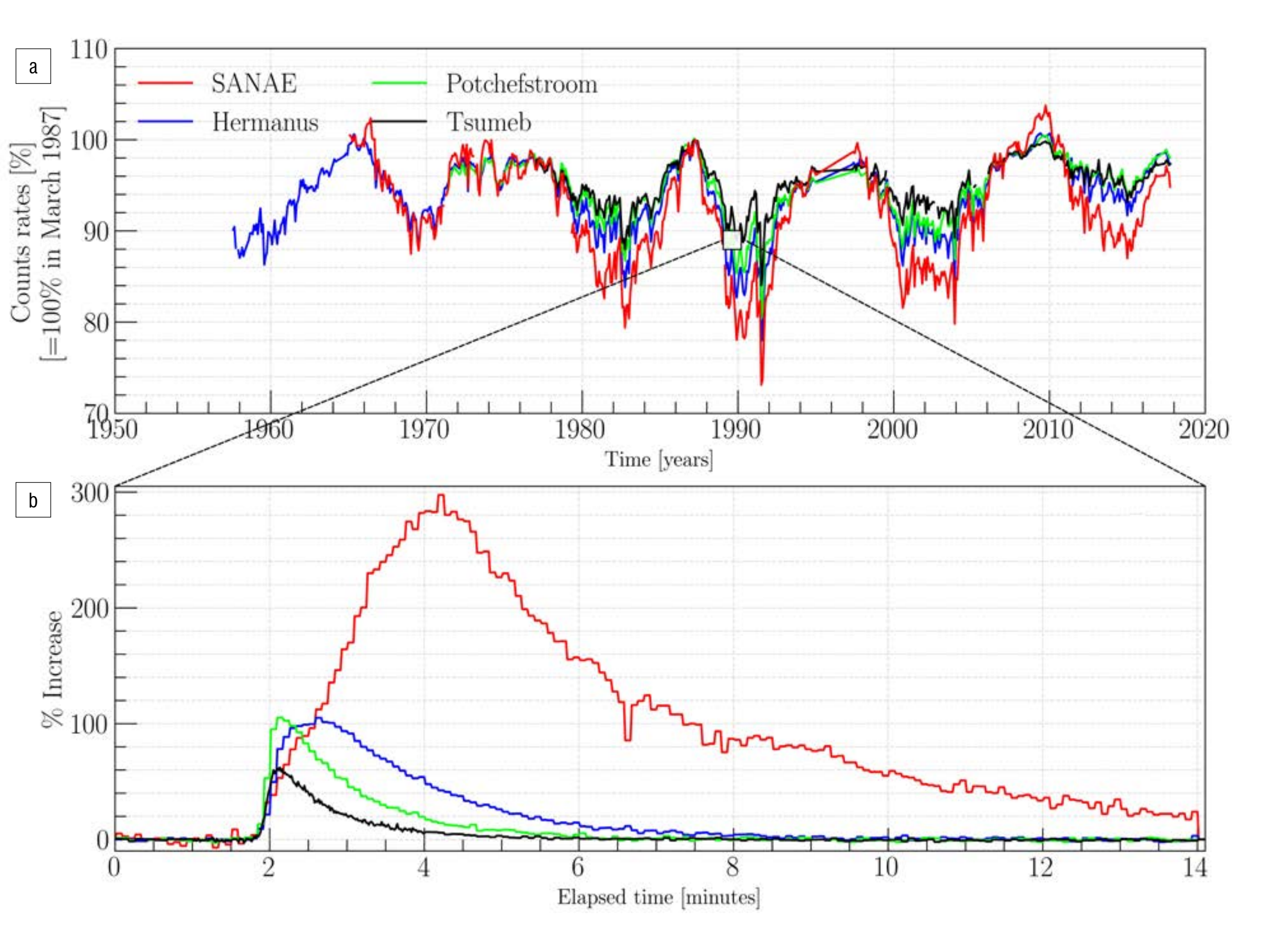


Figure 2: (a) Long-term and (b) short-term neutron monitor count rates observed by the four chosen neutron monitors. The data were taken from: http://natural-sciences.nwu.ac.za/neutron-monitor-data.

The High Altitude Radiation Monitor programme

The HARM programme is a student's project initiated by the NWU's Centre for Space Research. This programme offers opportunities for students to conduct technological and scientific experiments related to cosmic radiation in the atmosphere.

The HARM is a compact, battery-powered detector, which records cosmic radiation levels and is based on a Raspberry Pi micro-computer discussed below. The HARM prototype has been used to take measurements on board commercial aircrafts at aviation altitudes and also during high-altitude balloon launches. The balloon measurements allow the monitoring of the atmospheric radiation environment from ground level to high altitudes (in principle up to 40 km). These measurements of secondary and tertiary particles give information about the atmospheric radiation environment's evolution with altitude. The programme has been made available to students from other institutions through a collaboration with SANSA. Figure 3 shows students taking part in the HARM programme. Figure 3a shows Phillip Heita and Jacobus van den Berg from the NWU pictured with the first HARM prototype, while Figure 3b shows Portia Legodi from the University of Cape Town assembling an upgraded version of the HARM prototype, and Figure 3c shows Chris Thaganyana (pictured holding the inflator system nozzle and the balloon neck) from the University of the Western Cape taking part in the first HARM balloon launch.

The importance of this programme is to impart knowledge to the general public about the atmospheric radiation environment's evolution when flying, relative to altitude, geomagnetic position, solar activity, and flight duration. Apart from that, this programme assists students in gaining experience in designing, testing, and assembling relatively simplistic instruments to make novel measurements.

Instrumentation

Figure 4 shows the instruments comprising the basic detector system. All the sensors and modules are self-contained and draw their power directly from the Raspberry Pi computer which is powered with a standard power bank. For this particular HARM set-up, the following components were used:

- The Raspberry Pi 3 Model B+ micro-computer. This compact computer board offers endless opportunities for student projects. It is easy to add different sensors to this board, which can be programmed relatively easy. For more details, see https://www.raspberrypi.org/.
- The Raspberry Pi Sense HAT. The sensor measures environmental parameters such as pressure, temperature, humidity, acceleration and the magnetic field magnitude. These parameters are sampled once per minute.
- The J305β GM counter. The tube is cylindrical and capable of sensing low-energy ionising radiation. However, the counter will not normally detect neutrons as these do not ionise the gas inside the tube. The tube is powered from the Raspberry Pi by a radiation board from cooking-hacks.com (https://tinyurl.com/ox76ha5). The board also discretises the GM counters' pulses, which is read by the Raspberry Pi.
- All components are housed in a 3D-printed housing.

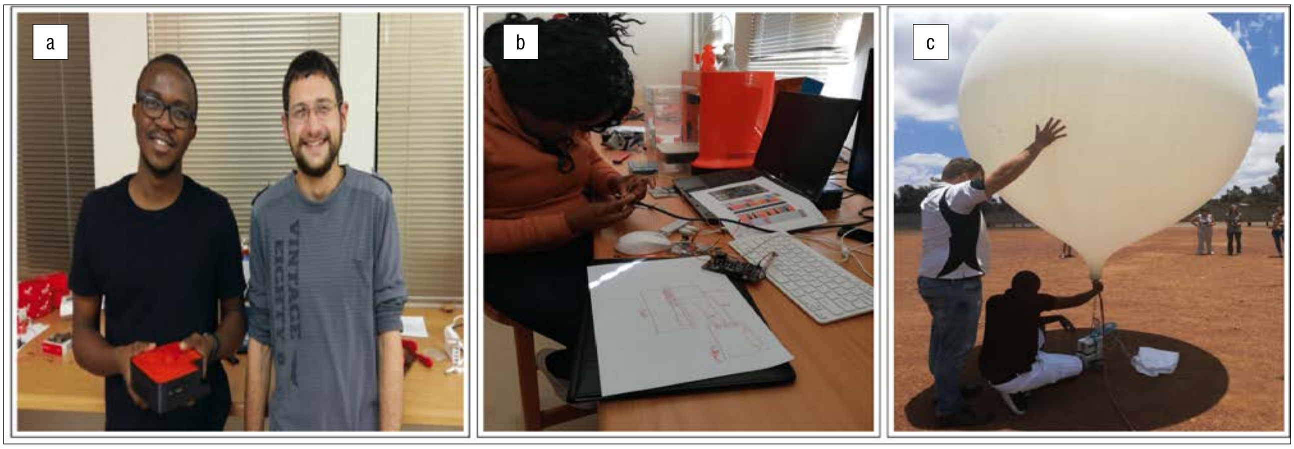


Figure 3: Students participating in the HARM programme. These activities occurred in the following order: (a) in June 2017 the first HARM prototype was assembled, (b) in June 2018 the prototype was upgraded and (c) in November 2018 the first HARM balloon launch took place.

The whole HARM set-up uses less power than a standard laptop, and does not radiate any high-frequency signals (all wireless connections, such as Bluetooth, are disabled). The detector can thus be safely used on board commercial flights. To fulfil several requirements to conduct more projects (including balloon launches), HARM has undergone several iterations resulting in different set-ups since the first version pictured in Figure 4. To reduce mass, the Raspberry Pi Model B+ was replaced with a smaller Raspberry Pi Zero W, while the Sense HAT was replaced with a BMP180 barometric pressure sensor. A real-time clock was added to keep the time, while a camera module was added for balloon launches. A camera was mounted outside a Styrofoam box to take pictures of the landscape and sky during the flight.

High-altitude balloon flight measurements

Aside from the scientific consideration, there is much more to be taken into consideration, such as applying for permission to launch the balloon from the Air Traffic and Navigation Services (ATNS) (https://www.atns.com/) – an entity of the South African Department of Transport. Prior to launching, the following equipment must be available: a meteorological rubber balloon, helium/hydrogen gas, an insulated payload box, detection system, location mapping tracker, and a parachute. All the instruments used are mounted inside, except the camera. The detector is tracked by placing a smartphone inside the box and using the standard tracking software of the phone.

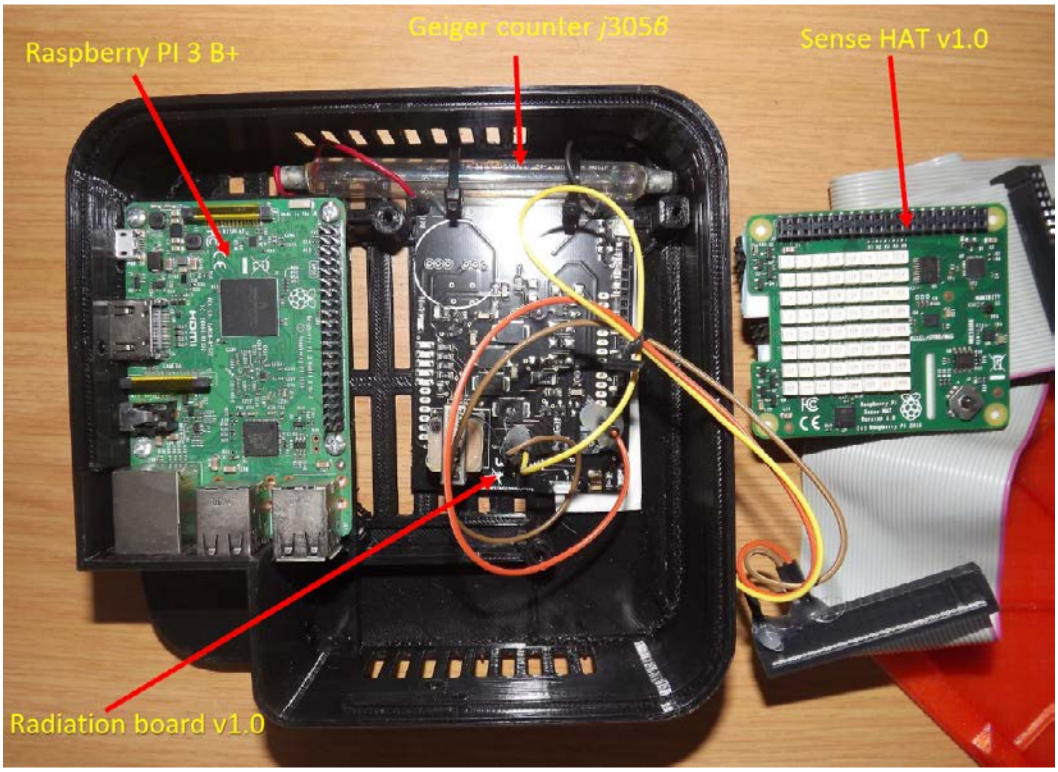


Figure 4: The HARM housing opened to reveal the detector electronics, showing the Raspberry Pi 3 Model B+, the GM tube and its control board, and the Sense HAT v1.0.

Launch site, burst altitude and landing site

To predict and estimate the flight path, burst altitude and landing site, this online website http://www.habhub.org can be useful. This process must be done a couple of days before the launch day. Once the launch site is chosen and the payload assembled, the ATNS can be contacted to apply for permission to launch the balloon.

For the recent balloon launch, a 1-kg natural rubber balloon was used. It was inflated with just enough helium gas to lead to a neck lift of about 1 kg. It was inflated until the payload and weights were lifted. With a payload ranging from about 1 kg to about 1.5 kg, the balloon was expected to produce an ascent rate ranging from about 5.5 m/s to 6.0 m/s. The balloon rises unguided and after reaching the burst altitude, the balloon bursts, and the payload returns to Earth with the help of a parachute of about 2.5 m in diameter. The descent rate of the balloon was estimated to range between about 7.5 m/s and 8.0 m/s as the payload approached the ground. The total duration of the flight was expected to be less than 5 h, depending on the wind patterns and weather conditions from the launch site to the burst altitude and back to the ground.

Figure 5 shows a summary of the high-altitude balloon launch. Figure 5a shows a balloon being prepared for launch, while Figure 5b shows a snapshot of the sky and the Earth horizon. Figure 5c shows the remains of a high-altitude rubber balloon just after it burst. Figure 5d shows the retrieved payload, at a distance from the launch site further than the predicted landing location, in an empty field.

The Model for Atmospheric Ionising Radiation Effects

The Model for Atmospheric Ionising Radiation Effects (MAIRE) from the RadMod Research Group was used in this work. MAIRE is a parameterised model based on particle transport calculations in the atmosphere using the FLUKA Monte Carlo code. The model is capable of calculating particle fluxes of different species using an incident GCR or SEP input spectrum at least up to an altitude of 100 km above sea level. These fluxes can be calculated for any desired date, geographic/geomagnetic flight position and solar activity. The simulated fluxes can then be used as an input parameter to calculate dose rates (in μSv/h) and accumulated doses (in μSv) for commercial aircrafts or for stratospheric radiation balloon measurements. The other calculation that can be taken into consideration is the possibility of a GLE occurring during the flight. These calculated values can then be used to compare with measurements obtained using a dosimeter. Currently, the model can be accessed only through the website www.radmod.co.uk/maire, due to the unavailability of its source code.

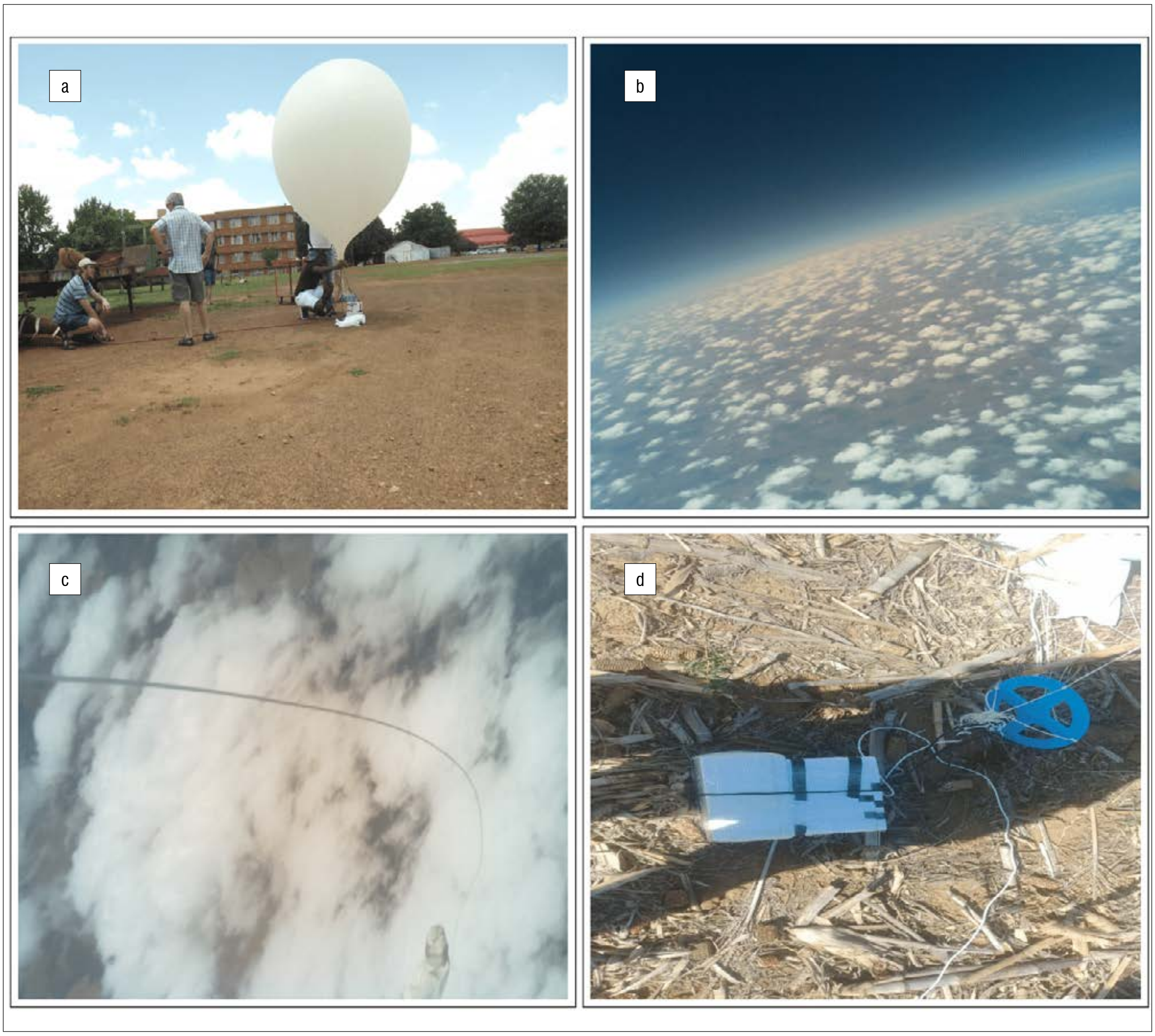


Figure 5: High-altitude balloon launch: (a) launch site, (b) Earth horizon, (c) the burst balloon captured during the descent, and (d) landing site.

Results and discussion

Measurements on board an aircraft

The measurements obtained during the first commercial flight of HARM are discussed below. The measured count rates are compared with the model calculations for the same date. Figure 6 shows the measured and calculated results.

In Figure 6a, the measured quantities are shown as a function of the elapsed time from departure. The count rates on board just before take-off and just before landing were not recorded as the device was switched off (see grey shaded areas) as requested by the aircraft flight personnel. The scattered error bar for each measurement was calculated as a Poissonian error. A 2-min moving average of the data (red solid line) was taken to show a smooth data trend. However, despite the shielding provided by the atmosphere, the HARM count rates rapidly increased above the background cosmic ray levels at ground level after take-off and levelled off at the cruising altitude recording count rate of between 25 per minute and 35 per minute, until the descent started before landing. The average value of the measured count rate at the cruise altitude was about 29 per minute.

Ideally, the first step of comparing the count rate measurements with model calculations is to use data sets acquired for the same date and time. These measured data sets must include information on the geographic flight position as well as corresponding measurements of at least one dosimetric parameter, e.g. the effective dose or ambient dose. However, because the HARM is unable to directly measure any dosimetric quantities (a GM counter produces only a count rate which cannot be related to a dosimetric quantity in a mixed radiation field), the data and model were compared in a qualitative fashion only. From Figure 6, the model results (blue solid line) show a general agreement with the count rate data (red scattered data), as measured by HARM, within some uncertainties.

The individual dose limit for members of the general public is 1 mSv/year (or an average dose rate of about 0.115 μSv/h over a year). However, if this limit is exceeded in a working environment (e.g. for nuclear plant workers), workers are classified as radiation workers, whose average dose limit is set to be 20 mSv/year (or an average dose rate of about 2.289 μSv/h; http://www.nnr.co.za/). Using the estimated doses from the model, the average effective dose rate from O.R. Tambo International Airport to Hosea Kutako International Airport is about 2.335 μSv/h at about 10 km above sea level for a duration of about 80 min, while the

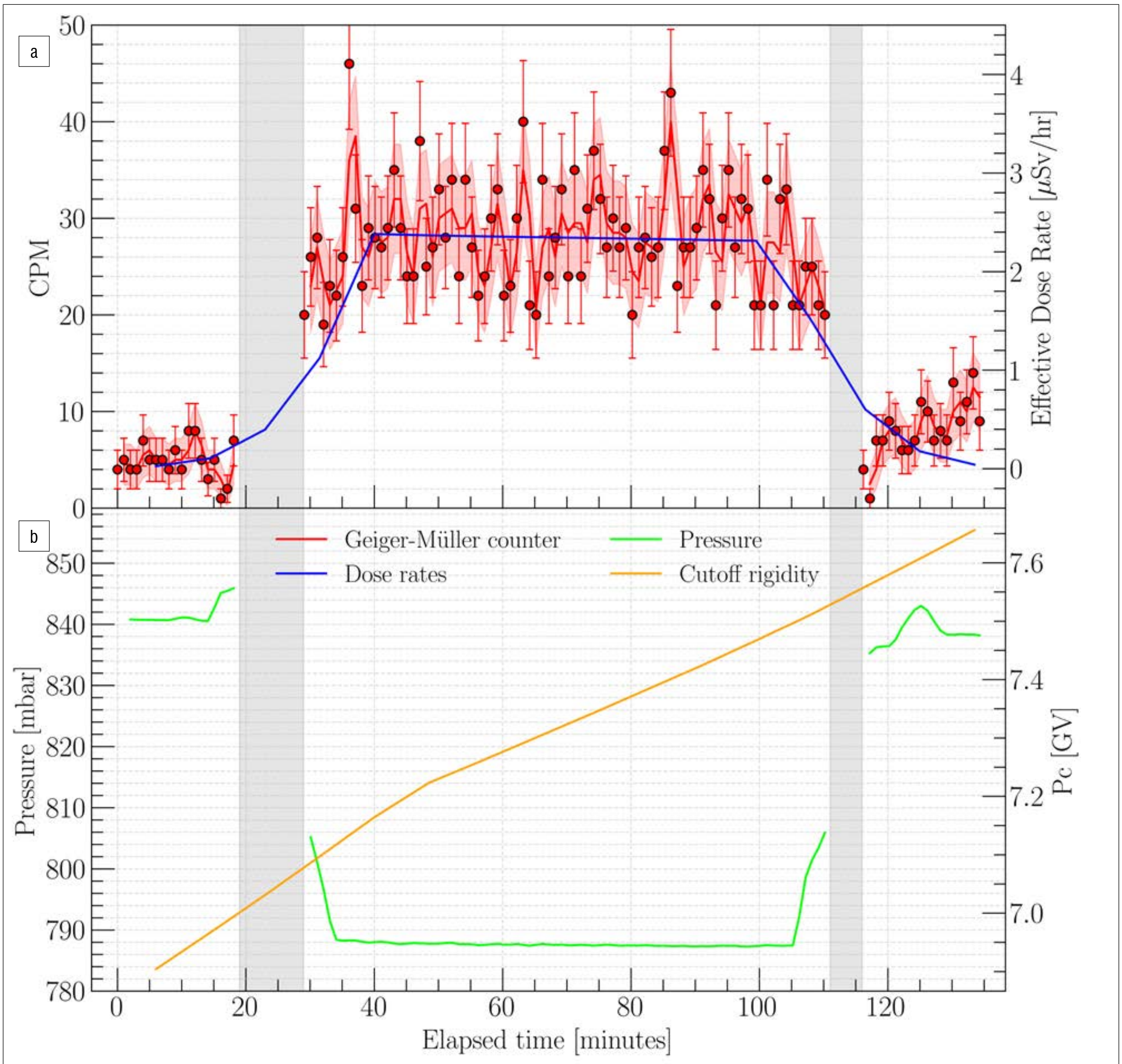


Figure 6: (a) The counts per minute (CPM; red colour) and model dose rates profile (blue colour). (b) Sensor HAT barometric pressure (green colour) and the Pc profile (orange colour).

HARM measured count rate is 29 per minute, on average. For an average member of the public, taking only several flights a year, the radiation exposure from flying is therefore negligible. However, if an aircrew member flies at least 285 short-haul flights per year (six such flights per week), using the same flight path and about the same flight time, they would receive a dose rate greater than 1 mSv/year – higher than the radiation exposure limit for the general public. This finding illustrates the need to define aircrew as radiation workers. These assumptions would be more realistic if the atmospheric conditions were the same, i.e. if there were no SEP events which might greatly enhance the atmospheric secondary radiation environment at cruising altitudes. In fact, aircrew in Germany have been classified as radiation workers as defined in the Radiation Protection Ordinance (StrlSchV) [28], while the German Federal Radiation Registry documents individual monthly effective doses for all aircrew members[29]. This is currently not the case in South Africa.

In Figure 6b, the aircraft cabin pressure profile is as expected when compared with the count rate profile: As the cabin pressure inside the aircraft changes due to the outside pressure changing, it directly corresponds to variations in altitude. As seen from Figure 6b, before take-off a small increase in pressure is expected when the cabin door closes and similarly when the cabin door opens after landing a small decrease is observed in the pressure profile. During take-off, because the detector was switched off for about 10 min, it appears that cabin pressure started rapidly decreasing and it levelled off at the cruising altitude, recording an average cabin pressure of about 787 mbar. Also shown in Figure 6b is the Pc profile from the MAIRE model calculation changing from 6.904 GV at O.R. Tambo International Airport to 7.656 GV at Hosea Kutako International Airport.

Measurements onboard a stratospheric balloon

Since the 1950s, an interest in observing cosmic ray particles through experiments with high-altitude balloons has been re-ignited[30] in the NM era. Among other discoveries made from these balloon launches, the most recent and notable is the disappearance of the RP-max by Hands and colleagues[31] using a low-cost portable radiation monitor. In South Africa, high altitude radiation dosimetry studies from ground level have not been documented before. However, the NWU (then known as Potchefstroom University for Christian Higher Education) conducted undocumented balloon launches for 20 years between 1958 and 1978 which characterised radiation levels with respect to ground level. Therefore, to address the unquantified and undocumented radiation levels, a balloon carrying the HARM instruments was launched at the NWU's Potchefstroom campus on 29 November 2018 after verifying in the laboratory that HARM's detection system was working and that all data were being received and recorded. The launch was successful and the measurements made by HARM are presented in Figures 7 and 8. The Pc at the launch site was 6.88 GV and at the landing site was 7.03 GV. These cut-off rigidities were calculated using the International Geomagnetic Reference Fields 2010 model.

From Figure 7, the profiles of the measurements, i.e. the count rate and atmospheric pressure and temperature, are plotted as a function of the elapsed time from launch. In both panels, the red and blue colours represent the ascend and descend phases, respectively, while the grey shaded area represents the burst altitude area. The red and blue vertical shaded areas in the ascend and descend phases represent the measurements at the aircraft cruising altitude between 10 km and 12 km. The count rate profile is shown in Figure 7a. The solid line represents a

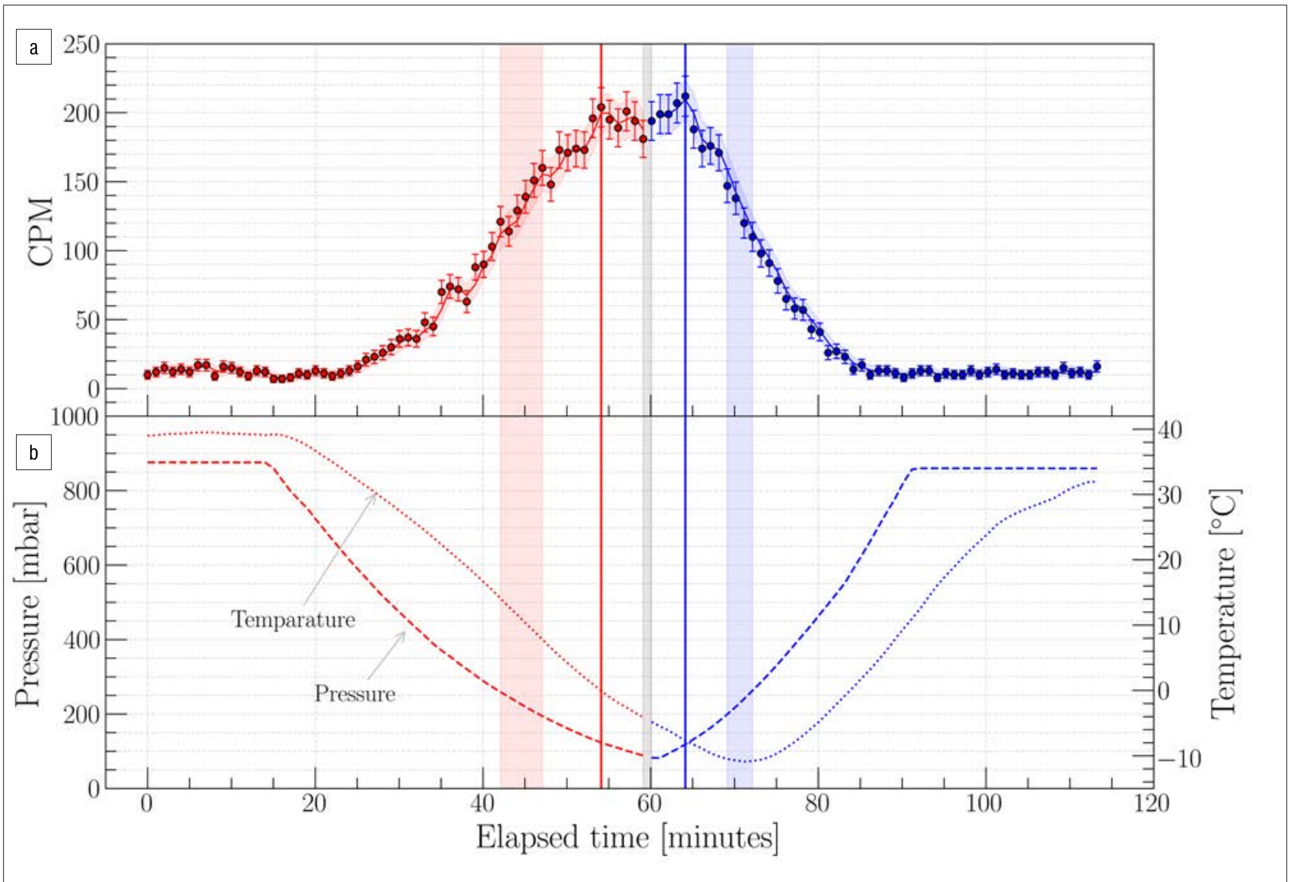


Figure 7: (a) The counts per minute (CPM) and (b) atmospheric pressure and temperature are shown as a function of the elapsed time from launch. The red and blue colours represent the ascend and descend phases, respectively, while the grey shaded area represents the burst altitude area. The red and blue vertical shaded areas in the ascend and descend phases represent the measurements at the aircraft cruising altitude between 10 km and 12 km. The blue and red thin vertical lines represent the RP-max before and after the burst altitude, respectively.

2-min running average, while actual measurements are shown in the original 1-min resolution as circles. At an ascent rate of 6.0 m/s from the launch site, it took the balloon about 27 min to reach the 10-km cruise altitude of commercial aircrafts. Clearly, from the shaded region in the ascent and descent phases, the balloon spent less time at the cruise altitude during the descent phase than it did in the ascension phase. Also seen from these data is what appears to be peaks, marked by the blue and red thin vertical lines during the ascent and descent phases. These peaks before and after the burst altitude are known as the RP-max. The profile of the measured atmospheric pressure and temperature, as a function of the elapsed time from launch, is shown in Figure 7b. The measured atmospheric pressure decreased as the count rate profile started increasing to high altitudes, as expected. At commercial flight altitudes, the pressure measured was between 200 mbar and 300 mbar, during both ascension and descension. The lowest recorded pressure and temperature was 82 mbar and -11 °C at the burst altitude (at the cruise altitude during the descent phase).

Figure 8 shows the correlation between the measured count rate and altitude. As pressure was measured using the BMP180 barometric pressure sensor and it is well known that pressure varies with altitude, the measured pressure from the sensor can be used to calculate an approximate measure of altitude. To approximate this altitude, an international barometric formula was used[32], which takes into account the average sea level pressure and the local pressure above sea level, and outputs an altitude value above sea level. The count rates were plotted together with the approximate altitude values. The shaded region between the 10-km and 12-km altitudes shows the count rate values at commercial flight cruise altitudes. At altitudes of about 14.7 km and 14.9 km, the count rate profile during ascent and descent phases shows clear peaks. These peaks seem to represent the RP-max.

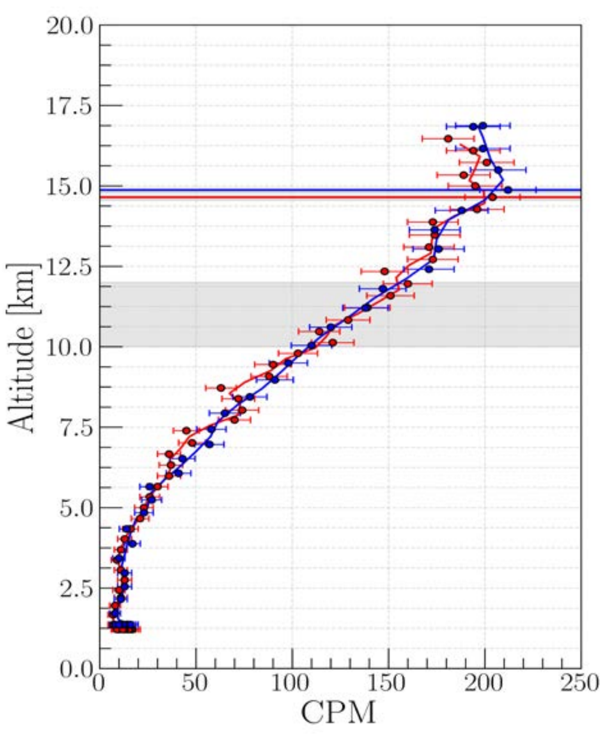


Figure 8: The counts per minute (CPM) profile of the balloon flight shown as a function of altitude. Red represents ascent, while blue represents descent. The grey shaded area between 10 km and 12 km represents the expected count rate measurement at the aircraft cruising altitude, while the horizontal lines at about 14.9 km and 14.7 km show the count rate measurements at the RP-max before and after burst altitude.

Conclusions

The biggest advantage of using the HARM prototype was that it is relatively low cost. Most of the components contained inside are off the shelf and only require knowledge of Python programming and rudimentary knowledge of sensor electronics. During balloon launches, most of the payloads used are recoverable, making it even more affordable and easy to relaunch using the same payload. The HARM programme has succeeded in measuring the mixed radiation field with respect to the geographic flight position during commercial aircraft flights and stratospheric balloon launches. These data sets are of interest to atmospheric scientists, space weather forecasters, commercial and private aircraft crew, and, most importantly, for public awareness. In principle, it is possible to obtain data every day and even during rare events like the SEP events from the ground up to the RP-max.

The HARM's first measurements of the RP-max and aviation radiation measurements on board a commercial flight in South Africa were presented. There is, however, room for improvement in these measurements. For one thing, the GM counter fails to measure the amount of energy being deposited by these particles upon contact with the tube. They also have a very poor response to neutrons. However, the inexpensive nature of the HARM detector and the relative ease of adding sensors to a Raspberry Pi micro-computer, make this an ideal student project. From the measurements obtained using HARM's GM counter it is not possible to calculate the dose rate. Therefore, the RPiRENA prototype, based on silicon semiconductor sensors, is being designed, in conjunction with researchers at the University of Kiel in Germany.[8,9]

As the only Space Weather Warning Centre in Africa, SANSA's appointment by ICAO as one of two regional centres to provide space weather services and warnings is a big boost for space science research in South Africa. As recommended by ICAO, SANSA must provide radiation estimates to the aviation sector, and the HARM programme can not only assist with this task, but also with raising public awareness regarding radiation levels at aviation altitudes.

Acknowledgements

We thank the students and researchers who participated in the HARM programme, from assembling the HARM prototypes to voluntarily carrying them on board aircrafts and taking measurements. We are especially grateful to Phillip Heita, Portia Legodi, Chris Thaganyana, Wessels Reitsma, and Gunther Drevin. We gratefully acknowledge the reviewers' valuable suggestions that improved the quality of the manuscript. Finally, we thank the North-West University and the South African National Space Agency for their collaborative effort in promoting high altitude radiation studies. We acknowledge financial support from the South African National Space Agency, the South African National Research Foundation, and the South African National Astrophysics and Space Science Programme.

Authors' contributions

G.M.M.: Conceptualisation, methodology, data collection, data analysis, writing – the initial draft, writing – revisions. R.D.T.S.: Conceptualisation, methodology, writing – revisions, student supervision, project leadership, funding acquisition. R.R.N. Methodology, writing – revisions, funding acquisition. P.J.v.d.B.: Conceptualisation, writing – revisions.

References

1. Ackermann M, Ajello M, Allafort A, Baldini L, Ballet J, Barbiellini G, et al. Detection of the characteristic pion-decay signature in supernova remnants. Science. 2013;339:6121. https://doi.org/10.1126/science.1231160
2. Ginzburg VL, Syrovatskii SI. Origin of cosmic rays. Geomagn Aeron. 1961;1:427.
3. Hillas AM. Cosmic rays: Recent progress and some current questions. arXiv. 2006. https://arxiv.org/pdf/astro-ph/0607109.pdf
4. Forbush SE. Three unusual cosmic-ray increases possibly due to charged particles from the Sun. Phys Rev. 1946;70:771–772. https://doi.org/10.1103/PhysRev.70.771

5. Van Allen JA. Observation of high intensity radiation by satellites 1958 alpha and gamma. J Jet Propuls. 1958;28:588–592. https://doi.org/10.2514/8.7396

6. Van Allen JA, Frank LA. Radiation around the Earth to a radial distance of 107,400 km. Nature. 1959;183:430–434. https://doi.org/10.1038/183430a0

7. US National Research Council Committee on the Biological Effects of Ionizing Radiation (BEIR V). Health effects of exposure to low levels of ionizing radiation: Beir V. Washington DC: US National Academies Press; 1990. https://doi.org/10.17226/1224

8. Möller T, Berger T, Böttcher S, Burmeister S, Ehresmann B, Heber B, et al. The calibration of the flight radiation environment detector (FRED). In: Proceedings of 33rd International Cosmic Ray Conference; 2013 July 2–9; Rio de Janeiro, Brazil. https://doi.org/10.13140/2.1.1179.9046

9. Ritter B, Maršálek K, Berger T, Burmeister S, Reitz G, Heber B. A small active dosimeter for applications in space. Nucl Instrum Meth B. 2014;748:61–69. https://doi.org/10.1016/j.nima.2014.02.030

10. Hellweg CE, Spitta LF, Koch K, Chishti AA, Henschenmacher B, Diegeler S, et al. The role of the nuclear factor κb pathway in the cellular response to low and high linear energy transfer radiation. Int J Mol Sci. 2018;19(8):2220. https://doi.org/10.3390/ijms19082220

11. Koch K. The role of nuclear factor kb in the cellular response to different radiation qualities [dissertation]. Cologne: University of Cologne; 2013

12. Hellweg CE, Baumstark-Khan C. Getting ready for the manned mission to Mars: The astronauts' risk from space radiation. Naturwissenschaften. 2007;94:517–526. https://doi.org/10.1007/s00114-006-0204-0

13. Wilson JW, Goldhagen P, Rafnsson V, Clem JM, De Angelis G, Friedberg W. Overview of atmospheric ionizing radiation (AIR) research: SST-present. Adv Space Res. 2003;32(1):3–16. https://doi.org/10.1016/S0273-1177(03)90364-4

14. Bazilevskaya GA, Krainev MB, Makhmutov VS. Effects of cosmic rays on the Earth's environment. J Atmos Sol-Terr Phys. 2000;62:1577–1586. https://doi.org/10.1016/S1364-6826(00)00113-9

15. Regener E, Pfotzer G. Intensity of the cosmic ultra-radiation in the stratosphere with the tube-counter. Nature. 1934;134:325. https://doi.org/10.1038/134325b0

16. Gaisser TK, Engel R, Resconi E. Cosmic rays and particle physics. Cambridge: Cambridge University Press; 2016. https://doi.org/10.1017/CBO9781139192194

17. Grieder PKF. Cosmic rays at Earth: Researcher's reference, manual and data book. Amsterdam: Elsevier; 2001.

18. South African National Space Agency (SANSA). SANSA Policy brief: Space weather impacts on aviation [document on the Internet]. c2016 [cited 2020 Feb 25]. Available from: https://tinyurl.com/y29ykjv7

19. Parker EN. Dynamics of the interplanetary gas and magnetic fields. Astrophys J. 1958;128:664. https://doi.org/10.1086/146579

20. Parker EN. The passage of energetic charged particles through interplanetary space. Planet Space Sci. 1965;13:9–49. https://doi.org/10.1016/0032-0633(65)90131-5

21. Alania MV, Modzelewska R, Wawrzynczak A. On the relationship of the 27-day variations of the solar wind velocity and galactic cosmic ray intensity in minimum epoch of solar activity. Solar Phys. 2011;270(2):629–641. https://doi.org/10.1007/s11207-011-9778-6

22. Bütikofer R. Cosmic ray particle transport in the Earth's magnetosphere. In: Malandraki O, Crosby N, editors. Solar particle radiation storms forecasting and analysis. Astrophysics Space Science Library vol. 444. Cham: Springer; 2018. p. 79–94. https://doi.org/10.1007/978-3-319-60051-2_5

23. Schlaepfer H. Cosmic rays (Spatium 11). Zürich: International Space Science Institute; 2003.

24. Simpson JA. The cosmic ray nucleonic component: The invention and scientific uses of the neutron monitor. Space Sci Rev. 2000;93(1):11–32. https://doi.org/10.1023/A:1026567706183

25. Clem J, Dorman L. Neutron monitor response functions. Space Sci Rev. 2000;93:335–359. https://doi.org/10.1023/A:1026508915269

26. Bazilevskaya GA. Solar cosmic rays in the near Earth space and the atmosphere. Adv Space Res. 2005;35:458–464. https://doi.org/10.1016/j.asr.2004.11.019

27. Lantos P. Radiation doses received on-board aeroplanes during large solar flares. Hvar Observatory Bulletin. 2003;27:171–178.

28. StrlSchV. Verordnung uber den Schutz vor Schaden durch ionisierende Strahlen [Ordinance on protection against damage caused by ionizing radiation. Koeln: International atomic energy agency; 2001. German.

29. Wollschläger D, Hammer GP, Schafft T, Dreger S, Blettner M, Zeeb H. Estimated radiation exposure of German commercial airline cabin crew in the years 1960–2003 modeled using dose registry data for 2004–2015. J Expo Sci Environ Epidemiol. 2018;28:275–280. https://doi.org/10.1038/jes.2017.21

30. Maurin D, Melot F, Taillet R. A database of charged cosmic rays. A & A. 2014;569:A32. https://doi.org/10.1051/0004-6361/201321344

31. Hands ADP, Ryden KA, Mertens CJ. The disappearance of the Pfotzer-Regener maximum in dose equivalent measurements in the stratosphere. Space Weather. 2016;14:776–785. https://doi.org/10.1002/2016SW001402

32. Grigorie TL, Dinca L, Corcau J, Grigorie O. Aircrafts' altitude measurement using pressure information: Barometric altitude and density altitude. WSEAS Trans Circuits Syst. 2010;9:1109–2734. https://dl.acm.org/doi/abs/10.5555/1865266.1865273

Check for updates

AUTHORS:
Temwa Moyo[1]
Sioux McKenna[2]

AFFILIATIONS:
[1]Centre for Higher Education Research, Teaching and Learning, Rhodes University, Makhanda, South Africa
[2]Centre for Postgraduate Studies, Rhodes University, Makhanda, South Africa

CORRESPONDENCE TO:
Temwa Moyo

EMAIL:
mtheto@gmail.com

DATES:
Received: 16 Jan. 2020
Revised: 11 May 2020
Accepted: 19 June 2020
Published: 29 Jan. 2021

HOW TO CITE:
Moyo T, McKenna S. Constraints on improving higher education teaching and learning through funding. S Afr J Sci. 2021;117(1/2), Art. #7807. https://doi.org/10.17159/sajs.2021/7807

ARTICLE INCLUDES:
☒ Peer review
☐ Supplementary material

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

EDITOR:
Jenni Case

KEYWORDS:
teaching development grant, historically disadvantaged institutions, historically advantaged institutions, University Capacity Development Grant, dropouts, graduation rates

FUNDING:
National Research Foundation (South Africa)

Constraints on improving higher education teaching and learning through funding

In the midst of massification, targeted funding has been used in various countries to address inefficiencies in teaching and learning. In South Africa, arguments have been made for significant investments to be made and the University Capacity Development Grant (UCDG) in particular is being used as a driver for improved outputs. Prior to its implementation in 2018, the UCDG comprised the Research Development Grant and the Teaching Development Grant. The Teaching Development Grant was intended to address low retention and throughput rates and ZAR5.5 billion was spent to this end over a 12-year period. The analysis presented here of all Teaching Development Grant budget plans and progress reports from 2007 to 2015 shows that the undifferentiated implementation of the Teaching Development Grant within a differentiated sector limited its potential for system-wide gains. Institutions without adequate resources tended to divert Teaching Development Grant funds to attend to backlogs rather than to address teaching and learning practices and such universities lost much of their allocation through the withholding of unspent funds. This blanket practice addressed the symptoms of underspending but not the structural, cultural and agential mechanisms that led to such under-expenditure. Uneven access to the limited teaching development expertise also impacted on the use of the grant. This call for a context-based approach to funding has been identified as a key success factor in grant interventions in both African and European universities. We recommend a sector-wide response in the form of a national body or plan for the benefit of all universities and investment in financial management enhancement.

Significance:

- The study contributes to a better understanding of how government funding interventions can achieve intended goals. The study calls for a more contextualised approach to funding and to greater collaboration across the sector to maximise limited capacity.

Introduction

With increasingly constrained funding available for higher education, performance-based funding regimes have been championed worldwide as a tool to steer universities towards improved quality and efficiency.[1-3] The University Capacity Development Programme, introduced in South Africa in 2018, was developed to promote staff development, curriculum development and student success in the system.[4] This programme includes the University Capacity Development Grant (UCDG) which is allocated to all public universities. The UCDG continues and extends the bold goals of the Teaching Development Grant (TDG) and the Research Development Grant (RDG) which have been in place, with variations in formula, since 2004. The collapsing of the RDG and TDG into the UCDG emerged in part in response to limited capacity for grant management at national and institutional levels and also as an attempt to steer the system into an integrated approach to institutional planning of research and teaching and learning activities, development and resource allocation.[5]

The RDG was introduced to build staff research capacity at all South African public universities and the TDG to support the enhancement of teaching and learning. The introduction of the TDG followed the publication of cohort studies of all first-time entering students conducted by the then Department of Education.[6] A subsequent in-depth analysis of cohort data was commissioned by the Council on Higher Education.[7] Such studies[6,8] indicated serious inefficiencies in the system, whereby only 33.9% of students studying in 3-year programmes were able to successfully complete their programmes within 4 years. The studies also looked at a longer period of 6 years, where only 47.1% of the studied cohort had graduated. Of further concern was that student performance was racially skewed, with African students, despite having the highest enrolment increases since the end of apartheid, being the poorest performers, with only 21.6% graduating in a 4-year period; 27.5% of mixed-race students graduated during this time, 32.1% of Indian students and 46.2% of white students.[6,9,10] A recent study[10] has shown a slight improvement in these output indicators but they remain low and racially skewed.

This social justice crisis is exacerbated by the low participation rate in higher education of youth 18 to 23 years of age, which in 2016 was around 19.1%.[11] Closer scrutiny shows that there are also significant discrepancies within participation rates with mixed-race and African students at 16.3% and 15.6%, respectively, and those for Indian and white students at 49.3% and 52.8%, respectively.[10,12,13] The call to address persisting inequities in access has been made by numerous stakeholders[14-17] and was also particularly observed in the 2015 and 2016 student-led protests which emerged in the form of the #feesmustfall and #rhodesmustfall movements[17,18]. These institutional protests called for an end to exclusionary cultures and structures at South African universities.[17,18] The protests highlighted that the university system was not serving South Africa's diverse society equally – be it through the fee structures, institutional traditions, curricula, or pedagogical approaches. Emerging from these protests was the 'assertion of the importance of concrete transformation and decolonisation in South African universities'[17]. The TDG thus became even more critical to achieve the transformation agenda.

The TDG was introduced in 2004 as part of the annual block grant that the South African Department of Higher Education and Training (DHET) pays to institutions and then as an earmarked grant from 2006. Earmarked grants are implemented to achieve particular policy goals and for which reporting on the use of such funds is required. Using earmarked funding to drive specific agendas is common around the world, and is variably known as project funding, set-asides, targeted funding or reserved funding.[19-21] The introduction of the TDG can be seen as an ambitious project that put teaching development at the forefront of the sector's conversation. TDG allocations to universities were differentiated in that amounts were based on the throughputs achieved by each university. This differentiated approach allowed universities achieving relatively low throughputs to be allocated larger amounts of TDG funding than those with higher throughputs.

In 2008, a review of the TDG was published, which argued for more focused use of the funds and recommended strengthened oversight on the part of government and better accountability by universities. Until the end of 2012, however, reporting on the use of funds remained through unaudited progress reports and accountability and monitoring structures were weak.[13,22] Despite the 2008 report and other calls in the literature for more focused interventions[23], it was not until 2013 that a policy to more explicitly guide and monitor usage was put in place. The delay can be attributed to capacity constraints at a national level[9,22] and to the restructuring of the DHET during that period. Calls for greater accountability regarding the use of state funds are an international phenomenon[24], and with ZAR5.5 billion spent on the TDG[25] over a 12-year period, and ongoing funds committed through the UCDG, it is imperative that we identify and address constraints on its potential to bring about improvements in teaching across the sector.

Methodological approach

It cannot be taken for granted that significant financial investments in education will lead to improved teaching and learning.[3,24,26,27] It is thus essential to make sense of how the TDG has been utilised and whether its implementation has resulted in system-wide gains. Here we draw from multiple data sources.[28] The 2008 TDG review report and TDG policy documents, such as the grant implementation criteria, were analysed alongside 343 documents relating to every public university in South Africa from 2004 to 2015 in the form of TDG budget plans, which detail how each university planned to use its allocated TDG funds and annual TDG progress reports, which contain financial reporting and narratives of TDG utilisation at each university.

Institutional and individual identities have been concealed. Institutional data references include mention of history, that is, historically disadvantaged institutions (HDIs) and historically advantaged institutions (HAIs) and merged institutions. Under apartheid, HAIs were designated for people of European descent and HDIs were designated for other population groups. Resource allocation and institutional autonomy were significantly skewed in favour of the HAI and apartheid legacies remain in evidence. The 'merged institutions' are those for which a merger process resulted in institutions that cannot be readily characterised as either HAI or HDI. Similar observations of resource differentiation have been made across sub-Saharan African higher education systems plagued by colonial legacies.[24]

Making sense of a phenomenon such as the TDG requires an analytical framework that allows for the investigation of complex social events.[27] Archer[29,30] posits that events and experiences in society emerge from the complex interplay of structural, cultural and agential mechanisms. She describes structures as social arrangements, resources or relations amongst social positions, as well as institutional and national arrangements, whilst culture refers to accepted, ingrained practices. Agency refers to the human ability to take action[31] and it is understood that the ability of agents to pursue their projects and interests will be enabled or constrained by pre-existing cultures and structures in which they find themselves. An understanding of these interactions provides explanatory power of how practices have emerged and persisted or been transformed in the sector. In order to understand events in the social world, Archer[29,30] argues that we need to enact analytical dualism so that, for the purposes of analysis, we separate out structure, culture and agency and identify the workings of each. In this study, we identified structures, such as resources, policies and offices that either enabled or constrained TDG implementation, certain cultures in the form of institutional practices that shaped TDG implementation and agential action that shaped TDG activities.

In the remainder of this article, we present three key issues that emerged in the data as constraints on the implementation of the TDG: these are financial constraints in the system, withdrawing of funds, and the uneven distribution of expertise both to lead and to implement teaching development initiatives.

Financial constraints in the system

South African universities operate in an environment profoundly affected by the socio-economic and politico-geographical realities of apartheid.[14,32] HDIs in particular struggle with shortages in areas such as basic operational funds, library resources, computer systems, accommodation for students, lecture venues and laboratories.[9,33-35] A grant specifically earmarked to address historical structural inefficiencies, the HDI Development Grant, was only effectively implemented in 2016.

The data show that this context significantly impacted on TDG implementation, with funds often being spent on infrastructure instead of on teaching development initiatives:

> *The TDG plays an important role for the capital acquisition plan and teaching equipment maintenance.* (Merged Institution 15)
>
> *In the absence of … laboratory facilities … we are obliged to outsource the teaching of some of the undergraduate courses to* [nearby historically advantaged university]. *The university thus requests funding for the enhancement of University-wide Teaching & Learning infrastructure.* (HDI 19)
>
> [Merged institution] *relies to a large extent on the annual earmarked funds from DHET to acquire and maintain capital equipment and educational technology.* (Merged Institution 17)

The TDG funds, particularly in the early years of implementation, were seen by many institutions to be for the purpose of upgrading and maintaining basic infrastructure.

> … [there is an] *unequal T&L* [teaching and learning] *infrastructure and unequal service provision on the different sites. In order to address this challenge the TDG was utilised mainly for providing equity on all learning sites in terms of computers in laboratories, laboratory equipment, audio visual technology, minimum standard in classrooms.* (Merged Institution 17)
>
> *It is estimated that an amount of R1.5 million would be required to bring equipment in lecture theatres on all campuses to an acceptable standard … In our view, teaching development funds in 2007 could legitimately be applied to this purpose.* (Merged Institution 8)

Chronic historical under-resourcing meant that the funds were used very loosely, and often in ways that had only tenuous links to the grant's purpose of improving teaching. Such approaches generally left teaching development untouched.[23] These practices were mainly evident in HDIs and on HDI campuses that had merged or been incorporated with HAIs.

Infrastructure is a necessary precondition for good teaching, so it is difficult to argue against the use of funds on such items. However, infrastructural requirements are not a sufficient condition for good teaching, nor are they the purpose of the grant as set out in the TDG documentation.

Some universities, in particular HAIs, had structural enablements to generate revenue from investments and third-stream income[14,15,36], which allowed them to augment the TDG funding and undertake larger-scale and more sustainable projects.

> …[the project] *would suffer a huge setback in their efforts at improving student retention and graduations, if the systems were abandoned as a result of diminished funding. In response, the university executive approved a special appropriation from the university's Main Fund to complement the DHET Grant…* (HAI 8)

While the data reveal financial augmentation of TDG projects at a number of HAIs, they also show that other institutions, mainly rural-based HDIs, were dependent on government funding with limited access to other sources of funding. This had particular implications given the misalignment of the government's financial year (from April to March) to that of universities' academic year (from January to December). Universities were required to submit progress reports at the end of April in each financial year to report on the utilisation of funds from 1 April of one year to 31 March of the next year. In reality, the stringent administration process necessary for the release of the funds, such as the assessments of progress reports (which took months), often meant that funds were not released until the third or even fourth quarter of the year. This had dire consequences for many institutions which were often then unable to implement their teaching development plans.

> *Sourcing experts to run workshops on particular areas of need … delayed due to late confirmation from DHET.* (HDI 5)

> *funds arrive in the 3rd month, when half of semester one has already been completed.* (Merged Institution 4)

Where there was financial capacity to advance the funds using internal monies, plans could be implemented from the beginning of the year while other institutions had only 4–6 months in which to do so. That the TDG implementation processes did not acknowledge these structural differences meant the potential gains of the TDG were limited at some universities.

The advancing of funds was impossible for some institutions with very limited internal reserves. Advancing of funds can also be seen to be an issue of how risk averse the financial management culture was within the university: while some universities were willing and able to advance internal funds for TDG projects, others that potentially had funds that could temporarily have been diverted to the projects were not willing to do so.

With the implementation of the UCDG in 2018, universities now receive funds at the beginning of the academic year. Despite this positive change in the national management of the earmarked grant, universities are still struggling to spend their approved budgets. At the end of the 2018/2019 financial year, about ZAR120 million of UCDG funds remained unspent. This suggests that underlying mechanisms shaping the inability of institutions to spend earmarked funds remain unaddressed.

Financial constraints were not only in relation to infrastructure and inability to augment or advance TDG monies, but also pertained to the expertise required to manage the funds. Problems related to the administration of the grant were often tied to the ways in which the institution was managed, providing what Archer[29,30] would term a structural constraint. There were ample examples in the data of institutions finding it difficult to track TDG expenditure or to ensure its use on approved items.

> *About R35 million was apparently utilised for other organisational business. The university management has promised to reverse this situation and also ensure the proper utilisation of earmarked funding.* (HDI 20)

> *At the moment, we are unable to provide a complete picture of how the above allocations have been expended by the various Departments … we request permission to submit the actual financials at the end of May.* (HDI 10)

> *There are delays in filling … positions due to very slow administration processes …* (HDI 20)

Overly bureaucratic administrative systems were reported in much of the data as a major constraint on the use of funding. Institutions with weak administrative systems were severely constrained in the implementation of the TDG. The failure of universities to utilise funds and to submit project plans and annual review processes on time suggests an institutional ethos that had been, at times, in crisis management mode.[37] In such cases, universities had no clear distinction in the roles of the governing body (Council) and leadership and management led by the Vice-Chancellor and Senate.[37,38]

Not much consideration was given to how the TDG could act as a mechanism to address issues of pedagogy and curriculum development at an institution-wide level. Many of the reports hinted at fairly ad-hoc project implementation and many proposals comprised multiple small projects to be run by individuals without an overarching institutional plan.

The institutional structures and cultures then curtailed the agency of those expected to manage and implement TDG projects.[39]

> *The University acknowledged that interventions introduced in the past had not been sufficiently effective because of a lack of ownership by the Faculties. The University has introduced the position of Executive Deans in the four Faculties with, amongst others, responsibility for all teaching and research management of the Faculty.* (HDI 23)

> *… our Departments were not fully aware that the grant existed and what it was for. … Implementation was … somewhat delayed and monitoring has not really taken place sufficiently.* (Merged Institution 4)

The combination of only being able to implement the planned TDG initiatives once the funding was received and having weak financial management systems with onerous bureaucratic requirements resulted in significant portions of the money remaining unspent and being withdrawn.

Withdrawing of unspent funds and resultant fluctuations in budgets

The process of withdrawing unspent TDG funds was introduced in 2013 and is a practice that has continued with the implementation of the UCDG. Given the loose ways in which the funds were used before 2013, it was indeed necessary to improve the DHET's oversight function.[40,41] This translated into the withdrawing of unspent funds through the withholding of the equivalent amount from the next year's grant.

This approach meant that while amounts had been specified for each university, the actual grant was largely dependent on successful expenditure in preceding years. In some cases, funds were withdrawn due to misuse, but much more common was that the initiatives for which the funds were intended were not implemented, leaving substantial funds unspent. The table below shows the funds that were withheld in 2014 and 2015 after the implementation of the 2013 TDG policy.

The HDIs were most affected by the more stringent monitoring processes given their incapacity to spend. The withheld funds were reinvested in collaborative projects which all universities could then access. In 2015, a portion of 7%, and in 2016 a portion of 10%, was top-sliced from each grant to contribute to the national collaborative funds.[40,41] In addition to these funds, the collaborative project pot was topped up with withheld funds. Institutions that were not able to spend their funds thus contributed more to the collaborative projects than those that had the capacity to spend their own funds.

Table 1: Funds withheld from universities in 2014 and 2015

2014	Amount unspent and withheld in following year
6 out of 7* HDIs	ZAR65 221 600
0 out of 10** HAIs	0
2 out of 5 Merged Institutions	ZAR29 619 200
2015	**Amount unspent and withheld in following year**
7 out of 7 (All) HDIs	ZAR39 852 342
4 out of 10 HAIs	ZAR18 968 268
3 out of 5 Merged Institutions	ZAR30 938 746

**In 2014 and 2015 there were a total of seven institutions designated as historically disadvantaged institutions (HDI). An eighth HDI was added in 2016 when one HDI was split into two universities in 2015.*

***There are arguably 10 historically advantaged institutions (HAIs). These include institutions that had HDI campuses incorporated into them or were merged with former HDI campuses.*

This approach better served institutions that had the capacity to spend as they first benefitted from their institution's TDG funds, and then from the collaborative projects. The structural inequalities were thus unintentionally reproduced through this process with the net distribution effect of the grant potentially being regressive and having the potential of widening the system's resource inequality gap. Similar observations have also been made in European university systems whereby 'those who perform well receive more money and thus have a relatively better position to perform in the next period, while those who performed less well receive less money and are thus in a weaker position for the future'[42]. This can lead to entrenchment of divisions.

While the amount allocated to each university was based on throughput statistics, the actual monies received were determined largely by institutional ability to manage budgets and implement projects. In some cases, the unspent funds constituted as much as 80% of the allocations.

Table 2: Institutions that had >80% of the 2014/2015 Teaching Development Grant allocations unspent and withdrawn

Institution	Amount unspent and withdrawn
HDI 20	ZAR26 676 800
Merged Institution 18	ZAR16 104 800
Merged Institution 1	ZAR15 610 400
HDI 23	ZAR11 151 200
HDI 9	ZAR8 519 200
HDI 19	ZAR7 124 000
HDI 5	ZAR5 961 600
HDI 10	ZAR5 788 800

Under apartheid, HDIs were allocated a prescribed annual budget which had to be spent within the calendar year. Such universities were neither permitted to invest funds nor to roll funding over to the next year. This historical practice of withdrawing unspent funds from HDIs constrained these universities from building financial management capacity and often led to a rush of spending towards the end of the year.[14,15] The withholding of TDG funds to ensure efficient spending now had the unintended consequence of echoing past practice.

The withdrawal of unspent funds does not address the underlying mechanisms constraining institutions' abilities to spend. Experience elsewhere in Africa has shown that the withholding of unspent funds encouraged unintended practices such as 'over-spending [in non-beneficial areas] and the misspending of funds'[26(p.46)] by highly centralised administrative systems.

In something of a vicious cycle, the risk in going ahead with planned projects prior to funding being received was much greater for HDIs which may not actually receive their full funds due to underspending in the previous years; but by not going ahead, the chances of funds being withheld increased. Those universities that could advance funds and were willing to take the risks associated with doing so were the same institutions that could spend their full funds and were thus unlikely to experience fluctuations.

The withholding of funds limited agential decisions as to how far universities could implement projects and this further shaped the type of projects that emerged. Sadly, this constraint was experienced particularly at those universities that largely depended on the TDG. The unpredictability of annual allocations posed challenges with regards to the continuity and full implementation of systematic longer-term interventions and, in some cases, this led to the suspension of such initiatives altogether. In as much as some[23,43] have argued for increased education investments for targeted interventions, this research highlights the constraints existing in the system that prevent universities from using the allocated funds appropriately. Money alone is not sufficient to bring about teaching development.

The uneven distribution of expertise in the higher education system

The unequal distribution of academic development expertise was a constraint on the successful implementation of teaching development programmes.[6] Jacob et al.[44] point out that shortage of expertise is arguably the greatest challenge facing African higher education systems. There is a need for a cadre of strong academic development structures and a critical mass of agents who can drive teaching development across the sector. However, the data indicate that some institutions experienced acute problems with attracting staff to implement TDG projects.[45] The ability to attract and retain staff is linked in part to apartheid differentiation[16,46] whereby HAIs are in urban settings while HDIs are in rural locations[9,47]. (The exceptions to this are the University of the Western Cape which is an HDI in an urban area and Rhodes University which is an HAI in a rural area).

The literature points to a number of problems in how teaching development work is undertaken. Quinn and Vorster[48] show that many attempts rely on problematic common-sense approaches which do not take theorised work on teaching and learning into account. An analysis of academic development centres at eight South African universities concluded that contextual features had a significant effect on the quality and uptake of professional development.[34]

The problems identified in the literature were evident in the data and were exacerbated by the nature of employment being offered. Generally, staff employed to drive TDG initiatives were employed on short-term contracts. The kind of expertise required to develop teaching is unlikely to be available under such conditions. Some universities were willing to take the financial risk of hiring people on 2- or 3-year contracts, secure in the knowledge that they would receive their full TDG allocation in those years. Other institutions were unable or unwilling to take such a financial risk and so hired people only as the annual TDG funding came in for what remained of that year. In many cases, such people, who would be responsible for working with academics on initiatives such as enhancing epistemological access by developing student writing or assisting academics in re-curriculation, were hired in administrative posts rather than as academics. This negatively impacted on who applied for such positions and what the minimum requirements were, limiting the credibility and the capacity of the incumbents.

Six universities reported in much detail on their inability to fill crucial posts, which prevented them from implementing most of their TDG projects. The inability of universities to fill vacant posts, be it due to the way the posts were structured, bureaucratic recruitment processes, or the inability to retain staff, jeopardised project implementation.

The thin distribution and stalled development of academic development expertise was evident in this study as those who were appointed were often not experts in the field having no structures to draw from in the implementation of teaching development strategic decisions.[49]This constrained them from 'articulating shared interests, organising for collective action, generating social movements and exercising corporate influence in decision-making as an empowered corporate agent would'[29(p.269)].

The individuals driving TDG projects within universities would, according to Archer[29,30], need to see this work as complimentary to their academic identities and capabilities. But even where this may have been the case, agency alone is insufficient. The exercise of agency to drive the TDG projects was conditioned by the nature of structures and cultures shaping the environment.[29,30] Many South African universities have managerial cultures with centralised power structures[14,34,45,50] which constrained the agency of those who might lead development projects.

The TDG task team report argued that given the uneven distribution of expertise across the system, a national initiative should be established to provide support for universities. Such a national initiative, the report advised, 'should be focused on training and developing staff at identified institutions in T&L enhancement, together with regular monitoring of the impact of such interventions'.[6,30] Shay[43] and Boughey[23] make arguments for focused system-level investments and interventions that look at curriculum structures and pedagogy to address sector-wide teaching output inefficiencies. While a national body has not been established to provide a base of such skills, the DHET did in 2015 launch the national collaborative programme, and although this programme is still in its infancy, it has already borne a number of inter-institutional projects.

Alongside the uneven teaching development expertise, the instability of management personnel to lead the TDG implementation also emerged in the data as a key constraint across the higher education system:

> *Due to challenges experienced in the finance department following the suspension and finally resignation by the then Director of Finance, the process of utilisation of the TDG was delayed.* (HDI 19)

Leibowitz et al.[34] indicate that strong leadership that contributes to cultures of professionalism is needed for teaching and learning enhancement. Corporate agents, agents with significant institutional power[29], are key to the success of teaching development work, thus the high turnover of such agents destabilised potentially enabling structures such as systems, processes and procedures of operations at universities.[17,51,52] Five universities have been placed under administration in the past 10 years[28,37], and two of these universities have been placed under administration more than once. This instability greatly constrained the use of the TDG for strengthening teaching in these institutions.

It should not be suggested that TDG implementation was without challenges in institutions where enabling mechanisms were in place. Despite some universities having implemented academic development work for more than 30 years, the uptake of this work was uneven. At times, individuals within institutions with less hierarchical cultures used notions of academic freedom to resist change and the institutional culture entailed engagement with teaching development on a voluntary basis, whereby those who dismissed the notion of teaching development could simply ignore such initiatives.[53]

> *Programme is moving ahead smoothly. Challenges to get buy in from all lecturers.* (HAI 2)

> *Faculties are [only] now beginning to be supportive of the integration of writing support into mainstream curricula.* (HAI 22)

Conclusion

We have distinguished between the *differentiated distribution* of the TDG, which was based on institutional throughput rates, and the *undifferentiated implementation* of this grant, through which universities were expected to use, manage and report on their funds in a uniform process. We highlight that, as much as some[43] have argued that increased investments are needed for improved teaching and learning, an increase in funding is not sufficient because structural and cultural constraints for appropriate implementation persist. Environmental factors shape institutional contexts and affect the possibilities of educational investments resulting in intended benefits. We argue that uniform implementation has constrained the potential to result in system-wide gains. The blanket implementation translated into some universities being better able to achieve gains from the TDG than others.

The practice of withdrawing unspent funds addressed the symptoms of underspending and not the structural, cultural and agential mechanisms that led to such under-expenditure. The blanket approach ignored the starkly differentiated nature of the system. This process did little to strengthen the positions of institutions that face large-scale inequities and constraints in their institutional structures. The HDIs which serve the most disadvantaged students arguably needed these funds the most given their historically based constraints for teaching and learning support, but they were also the most likely to have unspent funds withheld.[14,15] While a national academic development structure may bring problems in the provision of contextualised initiatives, it would seem that this might be a necessary process to more widely distribute the gains the TDG offers. Increased emphasis on collaborative grants that bring institutions together would also seem to be a useful mechanism to address constraints within specific universities. As a single public higher education system, the health of any one university relies heavily on the health of the sector as a whole.

Our study also shows the urgent need for improved grant management and financial processes in many universities. It would, however, be a mistake to interpret this as a need for more compliance structures. Instead, the data suggest this is a cultural issue, in Archer's[29,30] terms, requiring shared ownership of the teaching development project within the university. Management, administrative and academic staff need to have a commitment to reducing bureaucracy and efficiently implementing projects directed towards the development of teaching, and ultimately the improvement in student retention and throughput. This requires funding focused on ensuring improved financial management capacity with a concomitant reflection on institutional ethos.

The findings of this study that point to the problematic undifferentiated implementation of the TDG in a differentiated higher education sector are also applicable to other state interventions such as the Clinical Training Grant, Infrastructure and Efficiency Grant, the Foundation Provisioning Grant, the HDI-DG and the Veterinary Sciences Grant.

This article focuses on the constraints in the use of TDG funding and calls for them to be addressed, but this should not be read as a dismissal of the many successful interventions that have been implemented over the last dozen years. The history of the TDG traces numerous improvements in the management and use of the grant at both sector and institutional levels over time, and we argue that the recommendations from this study can ensure even better gains.

As the world moves into an uncertain financial future due to the economic impact of the COVID-19 pandemic, which is expected to impact universities' budgets as they restructure their operations to combat the spread and impact of the pandemic, so careful consideration of some of these findings is needed. It is possible that earmarked funds may be used to cover the many unexpected costs now arising. We would argue, however, that teaching development is now more urgent than ever and that national collaborations will be key to our response.

Acknowledgement

This research was undertaken within the South African National Research Foundation funded Institutional Differentiation project (NRF grant 97646).

Competing interests

We declare that there are no competing interests.

Authors' contributions

T.M.: Conceptualisation, methodology, data collection, sample analysis, validation, data curation, writing – the initial draft. writing – revisions. S.M.: Conceptualisation, writing – revisions, student supervision, project leadership, project management, funding acquisition.

References

1. Sampson A, Comer K. When the governmental tail wags the disciplinary dog: Some consequences of national funding policy on doctoral research in New Zealand. High Educ Res Dev. 2010;29(3):275–289. https://doi.org/10.1080/07294360903277372
2. Florich N. The politics of steering by numbers. Debating performance-based funding in Europe. Oslo: Rapport; 2008. http://hdl.handle.net/11250/280633
3. Strehl F, Reisinger S, Kalatschan M. Funding systems and their effects on higher education systems. OECD Education Working Paper 6. Paris: OECD Publishing; 2007. https://doi.org/10.1787/220244801417
4. South African Department of Higher Education and Training (DHET). Ministerial statement on higher education funding 2018/19 and 2019/20. Pretoria: DHET; 2017.
5. South African Department of Higher Education and Training (DHET). University Capacity Development Grant policy statement. Pretoria: DHET; 2017.
6. South African Department of Education (DOE). Report of the Task Team on Higher Education Teaching Development Grants. Pretoria: DOE; 2008.
7. South African Department of Education (DOE). White Paper 3: A programme for the transformation of higher education. Pretoria: DOE; 1997.
8. Scott I, Yeld N, Hendry J. A case for improving teaching and learning in South Africa in South African higher education. Higher Education Monitor. 2007;6:9–18.
9. South African Department of Higher Education and Training (DHET). Ministerial Committee on the Funding Review. Pretoria: DHET; 2013.
10. South African Department of Higher Education and Training (DHET). 2000 to 2008 First time entering undergraduate cohort studies for public higher education institutions. DHET: Pretoria; 2016.
11. South African Department of Higher Education and Training (DHET). Post school education and training monitor. Macro-indicator trends. Pretoria: DHET; 2019.
12. Badsha N, Cloete N. Higher education: Contribution for the NPC's National Development Plan. Cape Town: CHET; 2011.
13. South African Council on Higher Education (CHE). South African higher education reviewed – two decades of democracy. Pretoria: CHE; 2016.
14. Bozalek V, Boughey C. (Mis)framing higher education in South Africa. Social Policy Admin. 2012;46(6):688–703. https://doi.org/10.1111/j.1467-9515.2012.00863.x
15. Bunting I. The higher education landscape under apartheid. In: Cloete N, Maassen P, Fehnel R, Moja T, Gibbon T, Perold H, editors. Transformation in higher education. Higher Education Dynamics. Dordrecht: Springer; 2006. p. 35–52.
16. Muthama E, McKenna S. The contradictory conceptions of research in historically black universities. Perspect Educ. 2017;35(1):129–142. https://doi.org/10.18820/2519593x/pie.v35i1.10
17. Nyamnjoh A. The phenomenology of Rhodes Must Fall: Student activism and the experience of alienation at the University of Cape Town. Strat Rev South Afr. 2017;39(1):256–277.
18. Qwabe N. Protesting the Rhodes Statue at Oriel College. In: Chantikule R, Kwoba B, Nkopo A. editors. Rhodes Must Fall. The Struggle to Decolonise the Racist Heart of Empire. London: Zed Books; 2018. p. 6–16.
19. Geuna A, Martin R. University research evaluation and funding: An international comparison. Minerva. 2003(41):277–304 https://doi.org/10.1023/B:MINE.0000005155.70870.bd
20. Andoh H. Understanding the impact of grants at universities. University World News. 2019 June 14. Available from: https://www.universityworldnews.com/post.php?story=20190611124156418
21. Grove L. The effects of funding policies on academic research [doctoral thesis]. London: University College London; 2017.
22. South African Department of Higher of Education and Training (DHET). Ministerial statement on the use and management of the Teaching Development Grant. Pretoria: DHET; 2013.
23. Boughey C. Using the curriculum to enhance teaching and learning. S Afr J Sci. 2018;114(9–10), Art. #a0288. https://doi.org/10.17159/sajs.2018/a0288
24. Randriamahenintsoa E. Challenges and opportunities of higher education funding policies and programs in Madagascar. In: Teferra D, editor. Funding higher education in sub-Saharan Africa. London: Palgrave MacMillan; 2013. p. 147–183.
25. South African Department of Higher Education and Training (DOE). External report. Information on the state budget for universities. Pretoria: DOE; 2016.
26. Yigezu M. Funding higher education in Ethiopia: Modalities, challenges, opportunities and prospect. In: Teferra D, editor. Funding higher education in sub-Saharan Africa. London: Palgrave MacMillan; 2013. p. 38–70.
27. Ishengoma J. Funding higher education in Tanzania: Modalities, challenges, prospects and a proposal for new funding modalities. In: Teferra D, editor. Funding higher education in sub-Saharan Africa. London: Palgrave MacMillan; 2013. p. 214–246.
28. Moyo MT. An analysis of the implementation of the Teaching Development Grant in the South African higher education sector [PhD dissertation]. Grahamstown: Rhodes University; 2018.
29. Archer M. Realist social theory: The morphogenetic approach. New York: Cambridge University Press; 1995.
30. Archer M. Culture and agency. The place for culture in social theory. New York: Cambridge University Press; 1996.
31. Porpora DY. Morphogenesis and social change. In: Archer MS, editor. Social morphogenesis. Dordrecht: Springer; 2013. p. 25–37.
32. South African Council on Higher Education (CHE). Learning to teach in higher education in South Africa: An investigation into the influences of institutional context on the professional learning of academics in their roles as teachers. Higher Education Monitor. 2017(14):26–33..
33. Boughey C, McKenna S. A meta-analysis of teaching and learning at five historically disadvantaged universities. Pretoria: CHE; 2011.
34. Leibowitz B, Bozalek V, Van Schalkwyk S, Winberg C. Institutional matters: The professional development of academics as teachers in South African higher education. High Educ. .2015;69:315–330. https://doi.org/10.1007/s10734-014-9777-2
35. South African Department of Higher Education and Training (DHET). Report on the Ministerial Committee for the Review of the Provision of Student Housing at South African Universities. Pretoria: DHET; 2011.
36. Ntshoe I, De Villiers P. Steering the South African higher education sector towards transformation. Perspect Educ. 2008;26(4):17–27.
37. Parliamentary Monitoring Group (PMG). Universities under administration: Update by Department of Higher Education and Training, South African Parliamentary Committee on Higher Education [webpage on the Internet]. c2013 [cited 2020 Jan 13]. Available from: https://pmg.org.za/committee-meeting/16185
38. De la Ray C. Governance and management in higher education briefing paper prepared for the Second National Higher Education Transformation Summit [document on the Internet]. c2015 [cited 2020 Jul 27]. Available from: https://www.justice.gov.za/commissions/feeshet/docs/2015-HESummit-Annexure15.pdf
39. Colbeck C. How does a national research/education funding policy influence academics' professional identities and careers? Paper prepared for the Colloquium on International Policies and Practices for Academic Enquiry; 2007 April 20; Winchester, UK.
40. South African Department of Higher Education and Training (DHET). Ministerial statement on university funding. Pretoria: DHET; 2012.

41. South African Department of Higher Education and Training (DHET). Ministerial statement on university funding. Pretoria: DHET; 2015.

42. Pruvot E, Claeys-Kulik A, Estermann T. Designing strategies for efficient funding of universities in Europe. Brussels University European Association; 2013.

43. Shay S. Educational investment towards the ideal future: South Africa's strategic choices. S Afr J Sci. 2017;113(1/2), Art. #2016-0227. https://doi.org/10.17159/sajs.2017/20160227

44. Jacob W, Xiong W, Ye H. Professional development programmes at world class universities. Palgrave Commun. 2015;1(15002):1–27. http://dx.doi.org/10.1057/palcomms.2015.2

45. Garraway J, Winberg C. Reimagining futures of universities of technologies. Crit Stud Teach Learn. 2019;7:39–60. http://dx.10.14426/cristal.v7iSI.194

46. South African Department of Higher Education and Training (DHET). Report on the ministerial committee for the review of the provision of student housing at South African universities. Pretoria: DHET; 2011.

47. Buhlungu S. Speech on the occasion of inauguration of Vice-Chancellor of the University of Fort Hare [document on the Internet]. c2017 [cited 2020 Jul 27]. Available from: https://www.ufh.ac.za/files/Vice-Chancellor%20Inauguration%20speech.pdf

48. Quinn L, Vorster J. Theorising the pedagogy of a formal programme for university lecturers. In: Quinn L, editor. Re-imagining academic staff development: Spaces for disruption. Stellenbosch: Sun Media; 2012. p. 51–69.

49. Boughey C. Linking teaching and research: An alternative perspective? Teach High Educ. 2012;17(5):629–635. https://doi.org/10.1080/13562517.2012.725528

50. White K, Carvalho T, Riordan S. Gender, power and managerialism in universities. J High Educ Policy Manage. 2011;33(2):179–188. https://doi.org/10.1080/1360080X.2011.559631

51. D'Andrea V, Gosling D. Improving teaching and learning. A whole institutional approach. London: The Society for Research into Higher Education and Open University Press Imprint; 2005.

52. Luckett K. Working with 'necessary contradictions': A social realist meta-analysis of an academic development programme review. High Educ Res Develop. 2012;31(3):339–352. https://dx.doi.101080/07294360.2011.631518

53. McKenna S, Boughey C. Argumentative and trustworthy scholars: The construction of academic staff at research-intensive universities. Teach High Educ. 2014;19(7):825–834. http://dx.doi.10.1080/13562517.2014.934351

Check for updates

AUTHORS:
Tanja Verster[1]
Samistha Harcharan[2]
Lizette Bezuidenhout[3]
Bart Baesens[4]

AFFILIATIONS:
[1]Centre for Business Mathematics and Informatics, North-West University, Potchefstroom, South Africa
[2]Model Risk, First National Bank, Johannesburg, South Africa
[3]Model Risk, Absa, Johannesburg, South Africa
[4]Department of Decision Sciences and Information Management, KU Leuven University, Leuven, Belgium

CORRESPONDENCE TO:
Tanja Verster

EMAIL:
Tanja.Verster@nwu.ac.za

DATES:
Received: 11 Nov. 2019
Revised: 05 May 2020
Accepted: 23 June 2020
Published: 29 Jan. 2021

HOW TO CITE:
Verster T, Harcharan S, Bezuidenhout L, Baesens B. Predicting take-up of home loan offers using tree-based ensemble models: A South African case study. S Afr J Sci. 2021;117(1/2), Art. #7607. https://doi.org/10.17159/sajs.2021/7607

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:
pricing, retail credit risk, boosting, bagging, price elasticity

FUNDING:
South African Department of Science and Innovation

Predicting take-up of home loan offers using tree-based ensemble models: A South African case study

We investigated different take-up rates of home loans in cases in which banks offered different interest rates. If a bank can increase its take-up rates, it could possibly improve its market share. In this article, we explore empirical home loan price elasticity, the effect of loan-to-value on the responsiveness of home loan customers and whether it is possible to predict home loan take-up rates. We employed different regression models to predict take-up rates, and tree-based ensemble models (bagging and boosting) were found to outperform logistic regression models on a South African home loan data set. The outcome of the study is that the higher the interest rate offered, the lower the take-up rate (as was expected). In addition, the higher the loan-to-value offered, the higher the take-up rate (but to a much lesser extent than the interest rate). Models were constructed to estimate take-up rates, with various modelling techniques achieving validation Gini values of up to 46.7%. Banks could use these models to positively influence their market share and profitability.

Significance:

- We attempt to answer the question: What is the optimal offer that a bank could make to a home loan client to ensure that the bank meets the maximum profitability threshold while still taking risk into account? To answer this question, one of the first factors that needs to be understood is take-up rate. We present a case study – with real data from a South African bank – to illustrate that it is indeed possible to predict take-up rates using various modelling techniques.

Introduction

On a daily basis, banks receive home loan applications from potential customers. Depending on the customer's risk profile, affordability and other factors, the bank decides whether or not to offer a home loan to this customer. The risk profile and affordability dictate the interest rate and which loan amount (relative to the value of the house) will be offered. The take-up of these offered home loans influences the profit of a bank. If more customers take-up the offers, the profit can potentially increase (i.e. the bank's market share might increase) and if customers do not take-up these offers, the bank cannot potentially increase profit and market share. However, if more high-risk customers take up these offers, the bank might lose money due to customers defaulting. If low-risk customers decline these offers, the bank loses potential income. By understanding the factors that influence the take-up rates of home loans offered, the bank potentially benefits through increased market share and profits. In this paper, we build a model to predict the probability of take-up of home loans offered by focusing on interest rate[1] and loan-to-value (LTV)[2]. This take-up model relates to the responsiveness of a specific customer segment (based on, for example, the risk type of a customer) to a change in the quoted price. The 'price' of a home loan is the interest rate charged by a bank to the customer.

Banks improve their market share (and possibly also profitability) when they increase the take-up rate by offering different interest rates ('price') to different customers using risk-based pricing. To determine which interest rate to charge and for which customer, the bank needs to understand the risk levels and price elasticity of a customer; that is, how sensitive the customer is to interest rate changes. For example, at a price of 10%, a bank might sell the credit product (home loan) to 100 customers, yet at a price of 11% it would only sell to 90 customers. This emphasises the importance of understanding 'take-up probability' (also referred to as the 'price-response function').

The aim of this paper is threefold. Firstly, we investigate price elasticity on a South African home loan data set. To investigate the effect of *only* interest rate on take-up, we will build a logistic regression using only one covariate (i.e. interest rate). Secondly, we illustrate the effect of LTV on take-up rates in South Africa. Again, to illustrate this, a logistic regression is built using *only* LTV as the covariate. Lastly, we investigate whether it is possible to predict take-up rates of home loans offered by a bank using a combination of LTV and interest rates. Both logistic regression and tree-ensemble models were considered.

We focused primarily on the effect of interest rates and LTV on the take-up rates. Note that take-up rates are also influenced by other factors such as competitor offers, where another bank offers a home loan with more attractive terms (e.g. lower interest rate and higher LTV), which could hugely influence the take-up rate. Another factor is the turnaround time of an application, where a customer applies for a home loan at two different banks with similar loan terms. The bank that processes the application more swiftly is more likely to be accepted by the customer than the bank that takes longer to process the application.[1] These factors were not taken into account in this paper.

Interest rates and LTV

Our focus in this paper is to investigate how interest rates and LTV influence *take-up rates* of home loans. A fundamental quantity in the analysis of what price to set for any product, is the *price-response function* – how much the demand for a product varies as the price varies. This is the probability that a customer will take up the offer of a home loan. We will distinguish between take-up and non-take-up – the customer accepting (take-up) or not accepting (non-take-up) the home loan from the bank. As in Thomas[1], we will also use the terms 'take-up probability' and 'price-response function' interchangeably. The simplest price-response function is the linear function, but the more realistic price-response function is the logit function.[1] Within the retail credit environment, relatively little has been published about price elasticity, even though price elasticity is a well-known concept in other fields.

The effect of interest rates on take-up rates is also referred to as *price elasticity*. Phillips[3] outlines a number of reasons why the same product (e.g. a home loan) can be sold at different prices. Note that from the bank's viewpoint, banks typically 'price' for risk by charging a higher interest rate for higher-risk customers. From the customer's viewpoint, however, banks can also 'price' their loan product at different interest rates to increase market share (and possibly profitability).[4] Specifically, *price elasticity* can be seen as the willingness of a customer to pay for a product or service.[1,5] Pricing is a strategic tool[6] for acquiring new customers and retaining existing ones[7]. Limited studies of price elasticity have been done in emerging countries such as South Africa, for example the study on personal loans[5] and the study on micro-loans[8]. Very little research has been conducted on the price elasticity of home loans, both locally and internationally. In this paper, we investigate price elasticity on a specific home loan portfolio of a South African bank.

LTV is considered to be one of the most important factors in home loans lending – the higher the LTV, the higher the risk is from the bank's point of view.[2,9,10] The LTV ratio is a financial term used by lenders to express the ratio of a loan compared to the value of an asset purchased. In a paper by Otero-González, et al.[2], the default behaviour (risk) of home loan customers is explained using the LTV ratio. The influence of LTV on take-up rates is a 'chicken-and-egg' conundrum. The LTV offered to a customer will influence their take-up rate, but the LTV also influences the risk of the customer and their ability to repay the loan – the higher the LTV, the higher the risk of the bank losing money, as the sale of the property might not cover the home loan. On the other hand, the LTV offered to a customer is determined by the risk of the customer.[11] The bank will consider the risk of the customer to determine what LTV to offer, that is, a higher-risk customer will qualify for a lower LTV in order to prevent over-extending credit to the customer.

The same is true for interest rates. The *interest rate* offered to the customer influences take-up rates. However, the risk of a customer determines the interest rate offered to that customer, and the interest rate offered to the customer then influences the risk. The higher the interest rate, the higher the monthly repayment, which affects the affordability to a customer and thereby influences the risk of the customer.

Modelling take-up rates

The last aim of this paper is to predict take-up of home loans offered using logistic regression as well as tree-based ensemble models.

Logistic regression is commonly used to predict take-up rates.[5] Logistic regression has the advantages of being well known and relatively easy to explain, but sometimes has the disadvantage of potentially underperforming compared to more complex techniques.[11] One such complex technique is tree-based ensemble models, for example bagging and boosting.[12] Tree-based ensemble models are based on decision trees.

Decision trees, also more commonly known as classification and regression trees (CART), were developed in the early 1980s.[13] Decision-tree models have several advantages[14] – among others, they are easy to explain and can handle missing values. Disadvantages include their instability in the presence of different training data and the challenge of selecting the optimal size for a tree. Two ensemble models that were created to address these problems are bagging and boosting. We use these two ensemble algorithms in this paper.

Ensemble models are the product of building several similar models (e.g. decision trees) and combining their results in order to improve accuracy, reduce bias, reduce variance and provide robust models in the presence of new data.[14] These ensemble algorithms aim to improve accuracy and stability of classification and prediction models.[15] The main difference between these models is that the bagging model creates samples with replacement, whereas the boosting model creates samples without replacement at each iteration.[12] Disadvantages of model ensemble algorithms include the loss of interpretability and the loss of transparency of the model results.[15]

Bagging applies random sampling with replacement to create several samples. Each observation has the same chance to be drawn for each new sample. A decision tree is built for each sample and the final model output is created by combining (through averaging) the probabilities generated by each model iteration.[14]

Boosting performs weighted resampling to boost the accuracy of the model by focusing on observations that are more difficult to classify or predict. At the end of each iteration, the sampling weight is adjusted for each observation in relation to the accuracy of the model result. Correctly classified observations receive a lower sampling weight, and incorrectly classified observations receive a higher weight. Again, a decision tree is built for each sample and the probabilities generated by each model iteration are combined (averaged).[14]

In this paper, we compare logistic regression against tree-based ensemble models. As mentioned, tree-based ensemble models offer a more complex alternative to logistic regression with a possible advantage of outperforming logistic regression.[12]

In the process of determining how well a predictive modelling technique performs, the lift of the model is considered, where lift is defined as the ability of a model to distinguish between the two outcomes of the target variable (in this paper, take-up vs non-take-up). There are several ways to measure model lift[16]; in this paper, the *Gini coefficient* was chosen, similar to measures applied by Breed and Verster[17]. The Gini coefficient quantifies the ability of the model to differentiate between the two outcomes of the target variable.[16,18] The Gini coefficient is one of the most popular measures used in retail credit scoring.[1,19,20] It has the added advantage of being a single number between 0 and 1.[16]

Scenario setting

In the South African market, home loans are typically offered over a period of 20 to 30 years. If an application passes the credit vetting process (an application scorecard as well as affordability checks), an offer is made to the client detailing the loan amount and interest rate offered. Both the deposit required as well as the interest rate requested are a function of the estimated risk of the applicant and the type of finance required.

Ordinary home loans, building loans as well as top-up loans (a further advance on a home loan) are different types of loans offered in the retail sector.[21] The value of the property is obtained from a central automated valuation system accessed by all mortgage lenders.[22] Where an online valuation is not available, the property will be physically evaluated. Depending on the lender's risk appetite, a loan of between 60% and 110% of the property valuation will be offered to the applicant and is the LTV. The prime lending rate is the base rate that lenders use to make the offer, for example prime plus 2 or prime less 0.5. Mortgage loans are normally linked to interest rates and can fluctuate over the repayment period.[23] Fixed interest rates are normally only offered on short-term unsecured loans. The repurchase rate (repo rate) is determined by the South African Reserve Bank (Central Bank) Monetary Committee and is the rate at which the Central Bank will lend to the commercial banks of South Africa.[24] The prime rate is a direct function of the repo rate.

In all analyses, we subtract the repo rate from the interest rate to remove the effect of the fluctuations due to the fiscal policy that is reflected by the repo rate. This ensures that our analysis is not affected by the specific level of interest rate in South Africa. The analysis is done on the percentage above or below the repo interest rate. Note that as South Africa is a developing country, the repo rate fluctuates more frequently than it does in developed economies.

The sample consisted of 294 479 home loan approvals from one South African bank, with offers between January 2010 and July 2015. From these offers, 70% were taken up by the applicants for the varying LTVs and interest rates. The type of data available for each customer are:

- General demographic information gathered at application (e.g. income, gender, employment type).
- The application risk grade (the result of a home loan application scorecard resulting in five risk grades, with Risk Grade A being the lowest risk and Risk Grade E the highest risk).
- Information about the home loan offered (e.g. interest rate offered in terms of repo rate, the LTV, the term, type of loan i.e. building loan (B), further advance building loan (FAB), further advance ordinary loan (FAO), ordinary home loan (O); and an indicator as to whether the customer was new to this bank's home loan or not).

The risk grades are provided in Figure 1. The left side of Figure 1 indicates the lowest risk (Risk Grade A) and the right indicates the highest risk (Risk Grade E). The risk grade is usually derived from the results of a credit scorecard.[20,25]

Figure 2 shows the interest rates offered. Note that in Figure 2 the interest rate is adjusted by subtracting the repo rate.

The LTV offered is illustrated in Figure 3. A lower value of LTV indicates that the home loan value is less than the property value (i.e. 50% < LTV < 100%) and a higher value of LTV is where the home loan value is higher than the property value (i.e. LTV > 100%). An LTV higher than 100% can include additional costs (e.g. transfer cost), which is usually allowed for first-time buyers.

Price elasticity: Interest rate effect on take-up rates

To investigate the sensitivity of take-up to a change in the interest rate offered, a logistic regression was built. First, the data were split[26] into a training data set (70% or 205 802 observations) and a validation data set (30% or 88 677 observations), keeping the 30% non-take-up and 70% take-up rates in both data sets[18], in other words, stratified sampling[27]. The following data preparations were performed: subtract the repo rate from the interest rate; change class variables to numeric variables (using indicator functions); and scale certain variables (e.g. divide by 10 000).

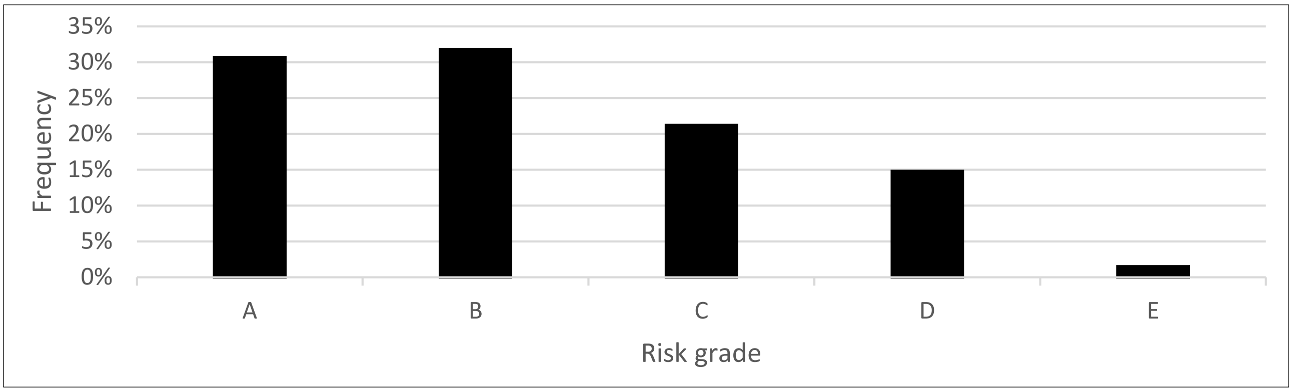


Figure 1: Risk grades of the home loans analysed (with Risk Grade A being the lowest risk and Risk Grade E the highest risk).

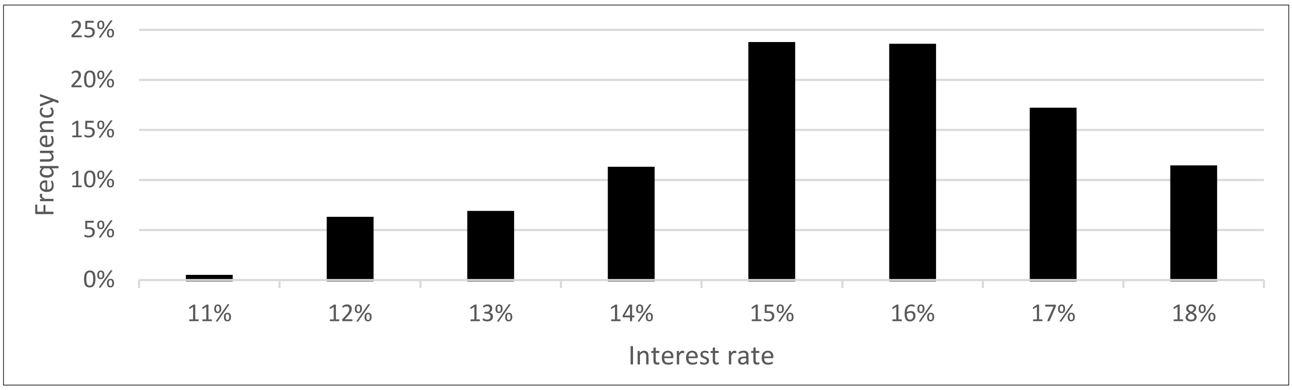


Figure 2: Interest rates offered for home loans analysed.

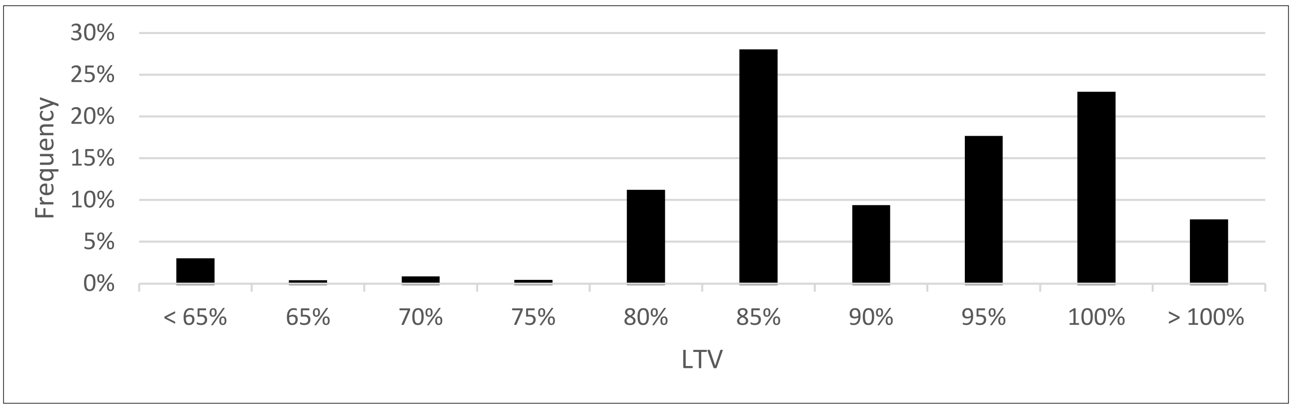


Figure 3: Loan-to-value (LTV) offered for home loans analysed.

A logistic regression model was built to predict a take-up rate given a certain interest rate (or LTV) offered. The probability of take-up is defined as the number of customers taking up a home loan divided by the number of customers who were offered a home loan. Note that the interest rate (and LTV) is an iterative process due to affordability (this relates to the chicken-and-egg conundrum). The resulting logistic regression is the price-response function. As mentioned before, a realistic price-response function is the logit function and therefore a logistic regression works very well in this context.

Using 5.5 years' of home loan data, a logistic regression was fitted on the interest rate:

$$p=\frac{1}{1+e^{-\{\theta\}}}$$

where $\theta=\beta_0+\beta_1 X_1\%$, and p is the probability of take-up and where X_1 is the recommended interest rate offered to the customer.

Due to confidentiality, the details of the estimates (β_0, β_1) are not given, but the logistic regression equation (p) is plotted in Figure 4. The Gini coefficient on the training data set was 0.316 and on the validation data set it was 0.314. The 95% confidence interval on the Gini coefficient on the validation data set was determined as (0.307; 0.322).

Figure 4 clearly shows that price elasticity exists in the home loans portfolio. The higher the interest rate offered, the lower the take-up rate. The take-up rates vary between 0% (very high interest rates) and 90% (very low interest rates offered). This illustrates the acceptance of loans that vary with the level of interest rate offered.

Interest rate effect on take-up rates: Different risk grades

We now investigate whether the price elasticity curves are the same for 'good' and 'bad' customers.

The logistic regression equation fitted to both the risk grade and interest rate shows that customers with medium to high risk – in other words, 'bad' customers – are less sensitive to rate changes than customers with low risk who are considered 'good' customers, as depicted in Figure 5. This shows that the take-up rate is higher for 'bad' customers versus 'good' customers for a given interest rate.

LTV effect on take-up rates

We now investigate how sensitive home loan customers are to the LTV offered. Again, note that the LTV is determined greatly by the risk of the customer as well as the size of the loan.

The following logistic regression was fitted:

$$p=\frac{1}{1+e^{-\{\theta\}}}$$

where $\theta=\beta_0+\beta_2\ (X_2)$ and p is the probability of take-up and where X_2 is the LTV calculated by the LTV rules in place at that specific time.

The Gini coefficient on the training data was 0.087 and on the validation data set was 0.093, with a 95% confidence interval of (0.084; 0.101). The Gini coefficient is much lower than the previous logistic regression developed for interest rates and we can reason that LTV has a much lower effect on take-up rates than does interest rate. Figure 6 illustrates the logistic regression equation fitted to LTV. We can see that the take-up rates now only vary between 59% and 74% (versus 0% to 90% in Figure 4).

Figure 6 shows the sensitivity of customers to LTV: the higher the LTV offered, the higher the take-up rate. This is not as prominent as the sensitivity of customers to interest rates when comparing Figure 4 with Figure 6. The reasoning behind this is that high-risk customers have limited options, as mentioned with interest rates. However, low-risk customers will either obtain a high LTV (resulting in take-up) or will be able to afford the deposit on a lower LTV loan (still resulting in take-up).

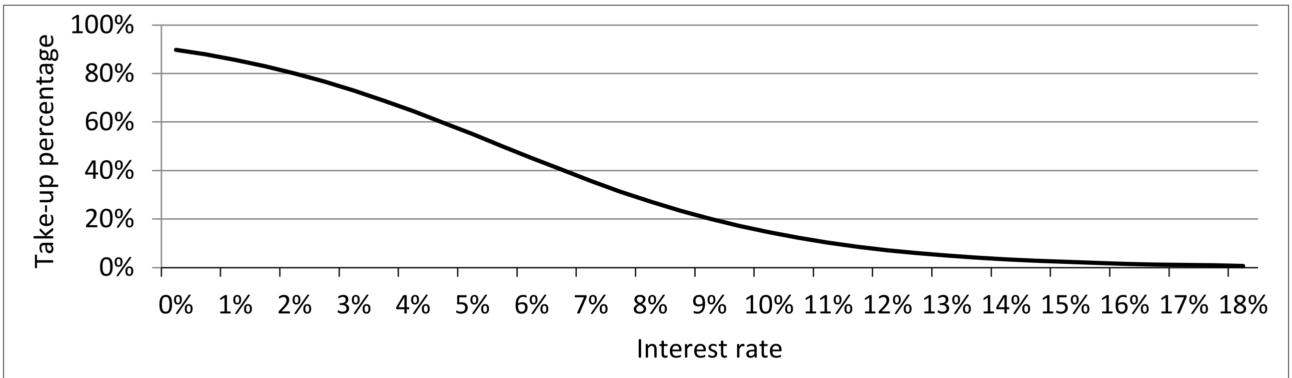


Figure 4: Probability of take-up given the interest rate offered (using logistic regression).

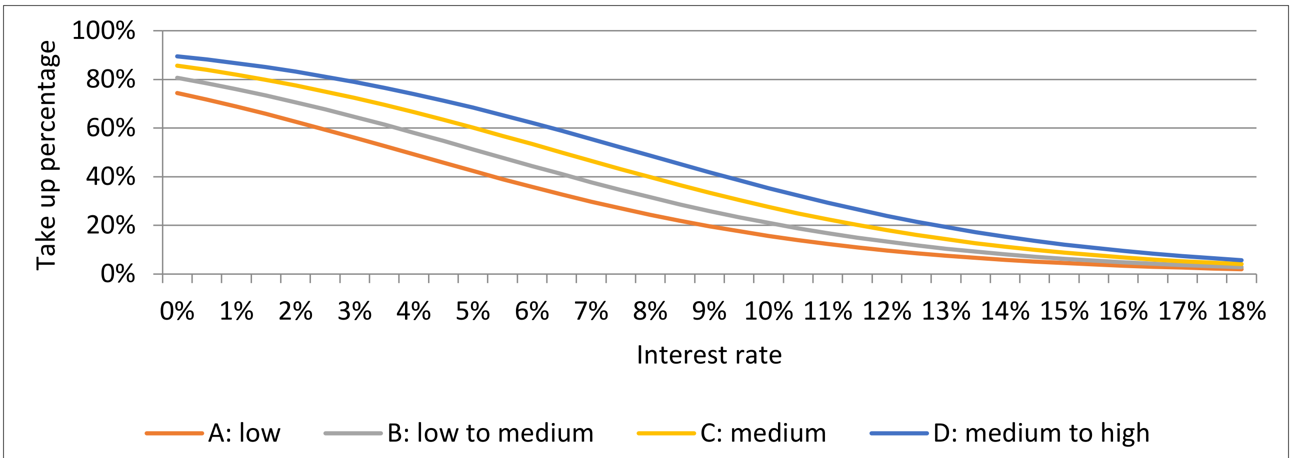


Figure 5: Probability of take-up given the interest rate offered (per risk grade).

Predicting take-up rates using logistic regression

Now we consider the combination of variables in logistic regression. All available variables were tested using a stepwise regression and six of the variables were significant at a 0.1% *p*-value. We fitted a logistic regression equation (again using 5.5 years' of home loan data):

$$p = \frac{1}{1+e^{-\{\theta\}}}$$

where

$$\theta = \beta_0 + \beta_1(X_1\%) + \beta_2 X_2 + \beta_3 I(X_3 = A) + \beta_4 I(X_3 = B) + \beta_5 I(X_3 = C) + \beta_6 I(X_3 = D) + \beta_7 I(X_4 = 1) + \beta_8 I(X_5 = B) + \beta_9 I(X_5 = FAB) + \beta_{10} I(X_5 = FAO) + \beta_{11} I(X_6 = 1),$$

and where:

p is the probability of take-up;

X_1 is the recommended interest rate (interest rate offered to client);

X_2 is the LTV calculated by the LTV rules in place at that specific time;

X_3 is the Applicant Risk Grade;

X_4 is 1 or 0 for employment type;

X_5 is the type of loan; and

X_6 is 1 for new to home loans and 0 otherwise.

The Gini coefficient on the training data was 0.410 and on the validation data set was 0.403, with a 95% confidence interval of (0.396; 0.412). Using the absolute value of the standardised estimates, the importance of the six variables is as follows: 1. interest rate; 2. employment; 3. LTV; 4. new to home loan; 5. risk grade and 6. type of loan. The logistic regression with more variables increased the previous Gini coefficient (where only interest rate was used) from 0.314 to 0.403 – an improvement of 28% on the validation Gini coefficient. The increase in the Gini coefficient on the training data could be due to the fact that more variables were used.[28] The fact that the Gini coefficient on the validation data set is similar to that on the training data shows that our model is generalising well and will be able to predict new cases.[29] Logistic regression is a relatively simple technique, and the question arose whether a more complex technique could further improve the model performance.

Predicting take-up rates using tree-based ensemble models

We investigate two tree-based ensemble models: boosting and bagging. Other techniques were also investigated, such as neural networks, rule induction and support vector machines, but were found to be less effective. A brief explanation of each of these techniques and the related results are provided in the supplementary material.

All models were built using the SAS Enterprise Miner software. SAS is a statistical software suite developed by the SAS Institute for data management, advanced analytics, multivariate analysis, business intelligence, criminal investigation and predictive analytics.[30] SAS Enterprise Miner is an advanced analytics data mining tool intended to help users quickly develop descriptive and predictive models through a streamlined data mining process.[30]

We have already mentioned that decision trees have several advantages and disadvantages and that ensemble models overcome these disadvantages while still maintaining the advantages. However, these ensemble models introduce their own disadvantages, namely the loss of interpretability and the transparency of model results. Bagging applies an unweighted resampling that uses random sampling with replacement, while boosting performs weighted resampling.

The bagging ensemble model resulted in a training Gini coefficient of 0.472 and a validation Gini coefficient of 0.467, with a 95% confidence interval of (0.460; 0.474). The boosting achieved similar results with a Gini coefficient on the training data set of 0.477 and on validation of 0.469, with a 95% confidence interval of (0.462; 0.477). From the Gini coefficient of 0.403 obtained previously using logistic regression, this improvement to 0.467 is a 16% increase on the validation Gini coefficient. The improvement of the Gini coefficient on the training data set could be due to the fact that we are using a more complex technique than logistic regression.[28] Note again the fact that the Gini coefficient on the validation data set is similar to the Gini coefficient on the training data, showing that the model did not overfit and in fact generalises well.[29]

Figure 7 shows the validation Gini with the 95% confidence interval. The 16% improvement using bagging or boosting (tree-based ensemble) on Gini is clear, but this comes at a disadvantage: the loss of interpretability and transparency. An overall decision needs to be made whether the improvement outweighs the loss of interpretability.

A summary of the abovementioned modelling techniques considered in this paper is given in Table 1, including the Gini results of both the training and validation data sets. It is clear that the tree-based ensemble models (bagging and boosting) outperformed the logistic regression.

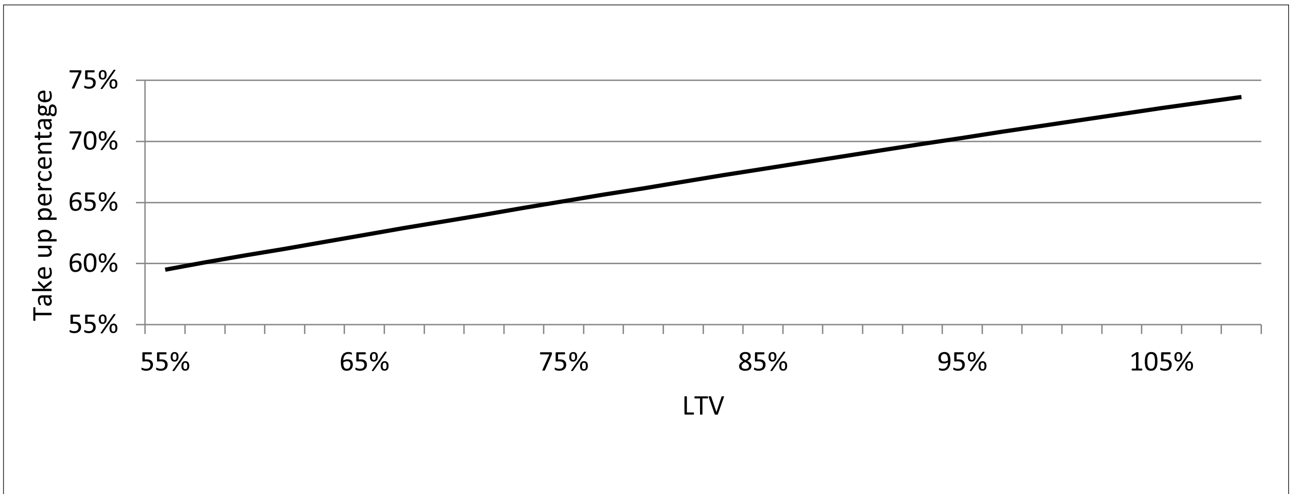


Figure 6: Probability of take-up given the loan-to-value (LTV) offered.

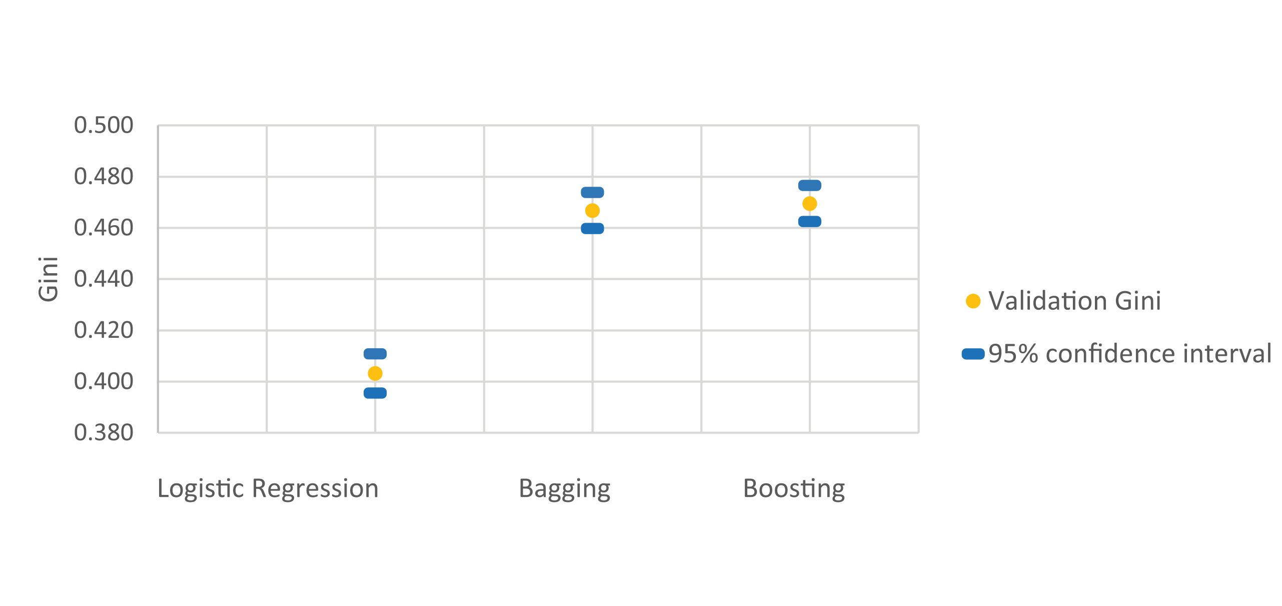


Figure 7: Validation Gini with 95% confidence interval.

Table 1: Gini results of different modelling techniques used

Modelling technique	Gini (training)	Gini (validation)	95% Confidence interval (Gini validation)
Logistic regression with interest rate	0.316	0.314	(0.307; 0.322)
Logistic regression with loan-to-value	0.087	0.093	(0.084; 0.101)
Logistic regression with all six variables	0.410	0.403	(0.396; 0.411)
Bagging with all six variables	0.472	0.467	(0.460; 0.474)
Boosting with all six variables	0.477	0.469	(0.462; 0.477)

Further investigation

The customers who did not take up the home loan offer were further investigated to determine whether they subsequently took up another home loan at another institution. This was attempted by using bureau data. Unfortunately, only 13% of these non-take-ups were matched on the bureau as taking up another home loan at another institution. There are many reasons for the low match, including identification numbers not matching (this could be due to a joint account).

If the customers did take up another home loan, we investigated whether they took up a more attractive home loan offer in terms of interest rate and LTV. A higher LTV and a lower interest rate were considered better offers, and vice versa.

Figure 8 shows the breakdown of the home loans offered at another institution. The results indicate that 22% moved due to a similar or worse deal, 11% moved due to a better (i.e. lower) interest rate, 48% moved due to a better (i.e. higher) LTV, and 19% moved due to a better interest rate and a better LTV.

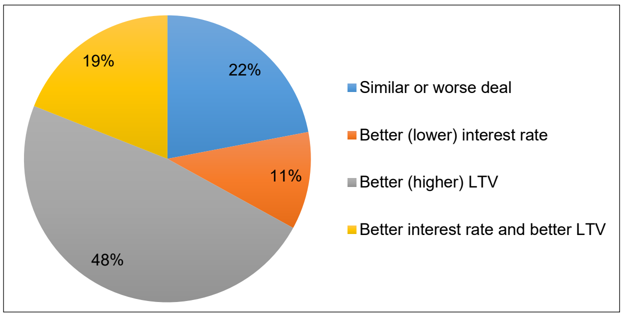


Figure 8: Breakdown of home loan offers taken up at another institution.

Conclusion and future research

The main contributions of this paper are threefold. Firstly, the effect of price elasticity in this specific South African's bank home loan database was illustrated. The higher the interest rate offered, the lower the take-up rate. Additionally, it was observed that high-risk customers are less sensitive to interest rate changes than are low-risk customers.

Secondly, we observed that home loan customers are sensitive to LTV: the higher the LTV offered, the higher the take-up rate (but not as sensitive as to interest rates offered). The 'chicken-and-egg' conundrum does pose some difficulty as the risk of a customer determines the LTV offered to the customer, and the LTV offered to the customer then influences the risk. Also, the LTV offered to the customer influences the take-up. A similar conundrum exists with interest rates.

Thirdly, models were built to predict the probability of take-up using home loan data over a 5.5-year period. Although logistic regression could predict take-up rates for home loan customers quite well, tree-based ensemble models can predict take-up rates more accurately (up to 16% improvement on validation Gini coefficients), but at a cost of interpretability.

The results of the bureau study indicate that 22% of customers moved to a home loan offered by another institution due to a similar or worse deal, 11% moved due to a better (i.e. lower) interest rate, 48% moved due to a better (i.e. higher) LTV, and 19% moved due to a better interest rate and a better LTV.

Many of the factors that influence take-up have not been captured into the models built in this paper, such as competitor offers, customer service, and so on. Unfortunately, it is very difficult to measure competitor offers. A future research topic could be to investigate the development of a factor that could reflect this influence.

With regard to adverse selection, lowering the interest rate could disproportionately draw high-risk customers.[31] This paper once again confirms that choosing the appropriate interest rate for a home loan is not as straightforward as it may seem.[4] In addition to general price sensitivity, adverse selection is an important characteristic in retail credit that is likely to have a significant impact on pricing.[32] In related literature, various definitions of adverse selection can be found. For instance, there is a distinction made between adverse selection on observable information and adverse selection on hidden information.[33] Phillips and Raffard[34] make the same distinction but refer to direct and indirect adverse selection. A further research idea can be to investigate whether adverse selection exists in the South African home loan market. According to Thomas[1], adverse selection needs to be taken into account as part of risk-based pricing as it influences the interaction between the quality of the customers and the probability of them taking up credit products.

In this paper, the Gini coefficient was chosen to measure model lift. A future research investigation could be to add other measures and compare the results when using these measures.

Note that this paper addresses a subset of the main question: What is the optimal offer that a bank could make to a home loan client to ensure the maximum profit for the bank while still taking risk into account? To fully answer this question, all the underlying factors need to be identified, then understood and, if possible, used to model these factors. This paper started with the take-up probability, which is one of the factors influencing the main question. Future research could be to expand on this to address other factors that influence the answer to this question.

Another possible future research idea could be to investigate the effect of the LTV policy rules on the models built that contained LTV as a variable. This specific bank had certain LTV policy rules in place at specific times. These LTV rules were based on, among other criteria, property values and risk grades. For example, if the property value was less than ZAR1.5 million and the applicant risk grade was C, the maximum LTV allowed would be 85%. These LTV rules regularly change at the bank and the influence of these policy rules could be researched.

Acknowledgements

This work is based on research supported in part by the South African Department of Science and Innovation (DSI). The grantholder acknowledges that opinions, findings and conclusions or recommendations expressed in any publication generated by DSI-supported research are those of the author(s) and that the DSI accepts no liability whatsoever in this regard.

Competing interests

We declare that there are no competing interests.

Authors' contributions

T.V. was responsible for conceptualisation; methodology; data analysis; writing the initial draft; writing revisions and project leadership. S.H. and L.B. were responsible for conceptualisation; data collection; data analysis; writing revisions; and project management. B.B. was responsible for critical reading of the manuscript and suggesting revisions.

References

1. Thomas LC. Consumer credit models: Pricing, profit and portfolios. New York: Oxford University Press; 2009. http://dx.doi.org/10.1093/acprof:oso/9780199232130.001.1
2. Otero-González L, Durán-Santomil P, Lado-Sestayo R, Vivel-Búa M. The impact of loan-to-value on the default rate of residential mortgage-backed securities. J Credit Risk. 2016;12(3):1–13. http://dx.doi.org/10.21314/JCR.2016.210
3. Phillips R. Pricing and revenue optimization. Stanford, CA: Stanford University Press; 2005.
4. Huang B, Thomas LC. Credit card pricing and the impact of adverse selection. Discussion Papers in Centre for Risk Research. Southampton: University of Southampton; 2009.
5. Terblanche SE, De la Rey T. Credit price optimisation within retail banking. Orion. 2014;30(2):85–102. http://dx.doi.org/10.5784/30-2-160
6. Nagle T, Holden R. The strategy and tactics of pricing: A guide to profitable decision making. Upper Saddle River, NJ: Prentice Hall; 2002.
7. Cross RG, Dixit A. Customer-centric pricing: The surprising secret for profitability. Bus Horiz. 2005;48:483–491. http://dx.doi.org/10.1016/j.bushor.2005.04.005
8. Karlan DS, Zinman J. Credit elasticities in less-developed economies: Implications for microfinance. Am Econ Rev. 2008;98(3):1040–1068. http://dx.doi.org/10.1257/aer.98.3.1040
9. Basel Committee on Banking Supervision. Basel II: International convergence of capital measurement and capital standards: A revised framework. Basel: Bank for International Settlements; 2006. Available from: https://www.bis.org/publ/bcbs128.pdf
10. Basel Committee on Banking Supervision. High-level summary of Basel III reforms. Basel: Bank for International Settlements; 2017. Available from: https://www.bis.org/bcbs/publ/d424_hlsummary.pdf
11. Engelman B, Rauhmeier R. The Basel II risk parameters: Estimation, validation, and stress testing. 2nd ed. Berlin: Springer; 2011.
12. Breiman L. Bagging predictors. Mach Learn. 1996;24:123–140.
13. Breiman L, Fredman J, Olsen R, Stone C. Classification and regression trees. Wadsworth, CA: Pacific Grove; 1984.
14. Maldonado M, Dean J, Czika W, Haller S. Leveraging ensemble models in SAS Enterprise Miner. Paper SAS1332014. Cary, NC: SAS Institute Inc.; 2014. Available from: https://support.sas.com/resources/papers/proceedings14/SAS133-2014.pdf
15. Schubert S. The power of the group processing facility in SAS Enterprise Miner. Paper SAS123-2010. Cary, NC: SAS Institute Inc.; 2010. Availabe from: https://support.sas.com/resources/papers/proceedings10/123-2010.pdf
16. Tevet D. Exploring model lift: Is your model worth implementing? Actuarial Rev. 2013;40(2):10–13.
17. Breed DG, Verster T. The benefits of segmentation: Evidence from a South African bank and other studies. S Afr J Sci. 2017;113(9/10), Art. #2016-0345. http://dx.doi.org/10.17159/sajs.2017/20160345
18. Verster T. Autobin: A predictive approach towards automatic binning using data splitting. S Afr Statist J. 2018;52(2):139–155. https://hdl.handle.net/10520/EJC-10ca0d9e8d
19. Anderson R. The credit scoring toolkit: Theory and practice for retail credit risk management and decision automation. New York: Oxford University Press; 2007.
20. Siddiqi N. Credit risk scorecards. Hoboken, NJ: John Wiley & Sons; 2006.
21. Property24. How banks assess loan applications [webpage on the internet]. c2011 [cited 2020 Feb 16]. Available from: https://www.property24.com/articles/how-banks-assess-loan-applications/14026
22. Lightstone Property [homepage on the internet]. c2020 [cited 2020 Feb 16]. Available from: https://lightstoneproperty.co.za/RiskAssessmentServ.aspx
23. Standard Bank. Home loans [webpage on the Internet]. c2020 [cited 2020 Feb 14]. Available from: https://www.standardbank.co.za/southafrica/personal/products-and-services/borrow-for-your-needs/home-loans/
24. South African Reserve Bank [homepage on the Internet]. c2019 [cited 2020 Feb 15]. Available from: www.resbank.co.za/Research/Rates/Pages/Rates-Home.aspx
25. Baesens B, Roesch D, Scheule H. Credit risk analytics: Measurement techniques, applications, and examples in SAS. Hoboken, NJ: Wiley; 2016.
26. Picard R, Berk K. Data splitting. Am Stat. 1990;44(2):140–147. https://www.jstor.org/stable/2684155
27. SAS Institute Inc. Applied analytics using SAS Enterprise Miner (SAS Institute course notes). Cary, NC: SAS Institute Inc.; 2015.

28. Friedman J, Hastie T, Tibshirani R. The elements of statistical learning. Berlin: Springer; 2001. https://doi.org/10.1007/978-0-387-21606-5_14

29. SAS Institute Inc. Predictive modelling using logistic regression (SAS Institute course notes). Cary, NC: SAS Institute Inc.; 2010.

30. SAS Institute [homepage on the Internet]. c2020 [cited 2020 Feb 15]. Available from: www.sas.com

31. Park S. Effects of price competition in the credit card industry. Econ Lett. 1997; 57:79–85. https://doi.org/10.1016/S0165-1765(97)81883-6

32. Stiglitz J, Weiss A. Credit rationing in markets with imperfect information. Am Econ Rev. 1981;71(3):393–410. https://www.jstor.org/stable/1802787

33. Ausubel LM. Adverse selection in the credit card market. c1999 [cited 2019 Jan 10]. Available from: http://www.ausubel.com/creditcard-papers/adverse.pdf

34. Phillips RL, Raffard R. Theory and empirical evidence for price-driven adverse selection in consumer lending. Paper presented at: 4th Credit Scoring Conference; 2009 Aug 26–28; Edinburgh, Scotland.

Performance assessment of four HIV self-test devices in South Africa: A cross-sectional study

AUTHORS:
Mohammed Majam[1]
Alex E. Fischer[1]
Naleni Rhagnath[1]
Vanessa Msolomba[1]
Willem D.F. Venter[1]
Laura Mazzola[2]
Samanta T. Lalla-Edward[1]

AFFILIATIONS:
[1]Ezintsha, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa
[2]Halteres Associates, San Francisco, California, USA

CORRESPONDENCE TO:
Alex Fischer

EMAIL:
fischermhealth@gmail.com

DATES:
Received: 17 Dec. 2019
Revised: 26 May 2020
Accepted: 25 Aug. 2020
Published: 29 Jan. 2021

HOW TO CITE:
Majam M, Fischer AE, Rhagnath N, Msolomba V, Venter WDF, Mazzola L, et al. Performance assessment of four HIV self-test devices in South Africa: A cross-sectional study. S Afr J Sci. 2021;117(1/2), Art. #7738. https://doi.org/10.17159/sajs.2021/7738

ARTICLE INCLUDES:
☒ Peer review
☒ Supplementary material

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

EDITOR:
Pascal Bessong

KEYWORDS:
HIV, HIVST, rapid diagnostic technology, sensitivity, specificity

FUNDING:
Tanya Shewchuk, BMGF

HIV self-testing (HIVST) has been introduced to supplement existing HIV testing methods to increase the number of people knowing their HIV status. Various HIVST kits have been developed; however, in many countries, their entry into the market is contingent on either being listed as World Health Organization (WHO) prequalified diagnostics/products or being approved by that country's health device regulator or both. In this cross-sectional study, we evaluated the usability, sensitivity and specificity of HIVSTs, as directed by the WHO prequalification literature. A boxed, sealed HIVST kit was provided to enrolled lay users with no further instruction, who then performed the test under observation. For each HIVST, a product-specific semi-structured checklist was used to calculate a usability index, while the sensitivity and specificity of each HIVST were calculated by comparing the HIVST results to the 'gold standard' – fourth-generation ELISA laboratory blood test. The average usability index was 97.1% (95.9–97.8%), while the average sensitivity and specificity were 98.2% (96.8–99.3%) and 99.8% (99.4–100.0%), respectively. We also diagnosed 507 (15.1%) HIV-positive participants from the general population. The average usability index, sensitivity and specificity were all comparatively high, and these results corroborate previous usability and performance studies from other regions. These results suggest HIVSTs are appropriate for the South African market and can assist manufacturers with readying their devices for final WHO prequalification evaluation.

Significance:

- This study has followed the WHO Technical Specification Series for the prequalification of HIV self-test devices, so the usability, sensitivity and specificity results may be used to inform the WHO prequalification process.
- The average usability index (97.1%), sensitivity (98.2%) and specificity (99.8%) were all very high, and these results support previous usability and performance studies from other regions, which suggest HIV self-tests are appropriate for WHO prequalification, and subsequently, the South African market.
- This study also diagnosed 507 (15.1%) HIV-positive participants from the general population – slightly higher than the national prevalence of 13.1%.

Introduction

The UNAIDS and the World Health Organization (WHO) 90–90–90 strategy released in 2015 has been adopted globally.[1] Despite significant progress made towards improving HIV testing rates in South Africa using the conventional, facility-based approach, it was still insufficient to reach the goal of testing 90% by 2020.[2] Inclusion of HIV self-testing (HIVST) in the South African strategy was considered to complement (by promoting use in populations who do not usually exhibit facility-based health seeking behaviour) and supplement (by providing a different option for HIV testing) existing methods while possibly improving HIV testing uptake, thereby facilitating target attainment.[3,4]

HIVST involves self-sampling of the user's oral fluid or blood specimen (dependent on the kit requirement), performing the HIV rapid diagnostic test (RDT), and then interpreting the result. The HIVST kits are intended to be used in a private setting, by a general population that encompasses a broad range of ages, education and literacy levels and nationalities. The benefit of HIVST includes immediate and confidential test results, and may encourage testing by groups who may otherwise avoid testing due to stigma, or the time and effort required for a clinic visit. HIVST can also promote more frequent testing, enable earlier diagnosis of HIV, may modify risk behaviours and may empower people to become more proactive and engaged in their health-care decisions.[5-7]

The first HIVST RDT approved for home use by the US Food and Drug Administration (FDA) was OraQuick ADVANCE Rapid HIV-1/2 Antibody Test in 2012[8] and since then, studies have continued to show the benefits of HIVST across several populations[5,9-11]. Based on this growing body of evidence, the WHO released guidelines for HIVST use in 2016, and strongly recommends HIVST as a way to supplement existing HIV testing services.[12] These guidelines recommend that only validated, WHO pre-qualified products should be used in public health programmes, and this position has also been adopted by the South African National Department of Health in their *National HIV Self Screening Guidelines 2018*.[13]

In order to validate products, the WHO Prequalification of In Vitro Diagnostics coordinated through the Department of Essential Medicines and Health Products has begun a prequalification process for HIVST to identify products which follow the best practices and standards set by international groups, including the International Medical Devices Regulatory Forum, the Global Harmonization Task Force, the US FDA and the European Regulatory Authorities.[14] In December 2017, the WHO released its Technical Specification Series for the prequalification process for HIV self-test devices. The WHO prequalification process includes a review of the device packaging, instructions for use, analytical and clinical performance data, as well as a manufacturing site inspection. Device manufacturers must also demonstrate that self-testing is supported by evidence from studies that explore usability and clinical performance, among a broad population of untrained intended users.[14]

The HIV Self-Testing Assessments and Research (HSTAR) programme at the Wits Reproductive Health and HIV Institute (Wits RHI) is a Bill and Melinda Gates Foundation funded programme to support HIVST developers looking to submit their device for prequalification and those seeking to enter the South African market, by independently providing data on HIVST usability (HSTAR001) and usability, performance and accuracy (HSTAR003) in the hands of untrained users.

The usability testing of seven prospective HIVST devices was recently completed with contrived results, as part of the HSTAR001 trial in Johannesburg, South Africa, and the usability index for each device was high, ranging from 84.2% to 97.6%.[15] Following a similar methodology, this study (HSTAR003) aimed to build on those results, and inform the WHO prequalification process by evaluating the usability of four HIVST candidates in clinical practice, with real-time results, instead of contrived ones. Additionally, the clinical performance and accuracy of these HIVSTs was investigated using sensitivity and specificity, by comparing results with the laboratory fourth-generation ELISA as the gold standard.

Methods

Study design

This cross-sectional study was implemented from March 2017 until November 2018, using the WHO prequalification published guidance. The HIVST devices were evaluated independently of the manufacturers and in series, to ensure no cross-contamination of assessments. To prevent participants from enrolling for more than one device, a fingerprint scanning Biometric Enrolment System was used.

HIVSTs

Four HIVST devices were assessed: three fingerstick whole blood devices and one oral fluid device. The three fingerstick devices were respectively produced by Biosure Ltd (United Kingdom), Biolytical Laboratories (INSTI) (Canada) and Chembio Diagnostic Systems (USA), while the oral fluid test was produced by Orasure Technologies (USA). Each HIVST device included the manufacturer's instructions for use (IFU) and other kit components, which were presented as intended for sale or distribution in South Africa. No additional job aids, demonstration or assistance were provided other than the manufacturer packaged materials.

Study participants

Convenience sampling was used to recruit adult participants from Wits RHI clinical trial sites in the inner city of Johannesburg. Included volunteers had to be at least 18 years old, had to be able to read English and to be first-time HIV self-testers with a self-reported unknown HIV status. Individuals were excluded if they had any prior experience with HIV self-testing or were health workers or lay counsellors who had performed HIV testing. Also excluded were participants who had received an experimental HIV vaccine or were taking HIV pre-exposure prophylaxis, persons known to be HIV positive or to have any extenuating condition (such as intoxication or acute sickness) which would interfere with the process.[15]

Using the WHO Prequalification Technical Specification Series document for guidance, a blended sample size of 900 participants was required for the usability assessment of each device. This sampling intended to blend high-risk and low-risk populations, and during training recruiters were made cognisant of recruiting equal gender participation, diverse age groupings and diverse education levels.[14]

Field procedures

All study procedures were conducted by a team of Good Clinical Practice trained researchers, and the self-testing followed the same procedures as HSTAR001[15] in that participants were handed a sealed test kit and they were provided with no further information about the device or test procedure. They were then requested to perform the test while being silently observed. The observer documented the process using a product-specific questionnaire. This was followed by a post-test interview.

Instead of being handed a contrived result to interpret, the participants' real self-test result was noted by the participant, then independently read and confirmed by a research nurse. In order to evaluate the performance and accuracy of the HIVST results, a 5-mL blood sample was drawn at the conclusion of each self-test, and a fourth-generation laboratory ELISA test (ABBOTT Laboratories, Chicago, USA) was performed within 24 h at the Wits Clinical Laboratory Services (a South African National Accreditation System (SANAS) approved, Good Clinical Laboratory Practice compliant facility). The ELISA laboratory test was used as the gold standard for the calculation of clinical sensitivity and specificity for each HIVST device.

HIV status was subsequently determined on site for all participants, irrespective of HIV status on the HIVST, using nurse-administered professional tests following the South African National Confirmatory Testing Algorithm.[13] Fingerstick samples were obtained using the Advanced Quality™ Rapid Anti-HIV 1&2 Test (RDT1) and the Abon™ HIV 1/2/O Tri-Line Human Immunodeficiency Virus Rapid Test Device (RDT2). If both the HIVST and RDT1 indicated a non-reactive/negative result, the participant was diagnosed as HIV negative. If one or both tests were reactive/positive, then the RDT2 test was performed. If both professional tests (RDT1, RDT2) were negative, then the participant was diagnosed as HIV negative. If both professional tests (RDT1, RDT2) were positive, then the participant was diagnosed as HIV positive and provided with a medical referral. In cases of discordant professional test results, the ELISA test was used for final diagnosis, and the participant was referred to a clinical site for the test results and follow-up.

Data collection

For the recently completed HSTAR001 usability assessment, the WHO prequalification literature was used to design, pilot test and implement a product-specific semi-structured questionnaire for data collection[15] which was also used in the current HSTAR003 study. The usability questionnaire comprised a HIVST process checklist guided by IFU steps, used to calculate usability index and a post-test interview that investigated the participants' competency, experiences and recommendations. For performance and accuracy evaluations, the ELISA laboratory test results were provided back to the research staff as an electronic copy within 24 h via email, and a hard copy was hand delivered within 7 days.

Data analysis

After data collection, field workers transcribed the questionnaire results into an MS Excel database. Quantitative data were analysed with descriptive statistics. Each batch of test kits went through a quality control check and 10% of all data entries were also checked by administrators for quality control.

Sensitivity and specificity were analysed to measure the performance and accuracy of each HIVST. Sensitivity refers to the ability of the HIVSTs to accurately detect truly positive tests, while specificity refers to the ability of the HIVSTs to correctly filter out truly negative test results. Both outcomes improve as they approach 100%, and their calculations are presented in Figure 1. The data supporting the results of this study are available upon request to the corresponding author.

Ethical considerations

The study was approved by the Human Research Ethics Committee of the University of the Witwatersrand (No. 161110). All participants signed an informed consent form and participants received a reimbursement for their participation. The manufacturer played no part in the study design, procedures or analysis of findings.

Sensitivity = [TP / (TP + FN)] x 100, where

- TP (true positive) is positive HIVST results, in agreement with positive ELISA laboratory results, and;
- FN (false negative) is negative HIVST results, discordant with positive ELISA laboratory results.

Specificity = [TN / (TN + FP)] x100, where

- TN (true negative) is negative HIVST results, in agreement with negative ELISA laboratory results, and;
- FP (false positive) is positive HIVST results, discordant with negative ELISA laboratory results.

Figure 1: Sensitivity and specificity calculations.

Results

Demographics

Table 1 presents the demographic data of participants who tested each HIVST; there was a diverse distribution of age groupings and education levels. The majority of participants were South Africans (3201/3600; 88.9%) under 35 years of age (2842/3600; 78.9%) and just over half of them (1944/3600; 54.0%) were men. The majority of participants had graduated secondary school (2056/3600; 57.1%) or attended tertiary school (1428/3600; 39.7%) while only 116/3600; 3.2%) had primary school or less. Only 853 (23.7%) were employed, while 2279 (63.3%) were unemployed and 467 (13.0%) were students.

Usability assessment

The four HIVSTs had an average usability index of 97.1% (95.9–98.8%) on their product-specific usability assessment (Table 2). The full usability indexes for each HIVST are available in Supplementary table 1. Despite the high usability, there were several spoiled tests (233/3600; 6.5%), in which critical errors prevented the test from producing a valid result. The majority of spoiled tests came from specimen collection errors (101/3600; 2.8%) or process errors (160/3600; 4.4%). A small number of spoiled tests were due to participants asking for assistance (7/3600 (0.2%) or quitting (12/3600; 0.3%). Four (0.1%) participant results were also deemed invalid due to defective kits, as they did not present a positive internal control line, even though the participants correctly completed all steps.

The process and collection errors that limited usability were specific to each device. Common errors across most fingerstick devices were due to incorrect lancing technique or lancet placement, resulting in insufficient blood available, failure to transfer the blood specimen to the device or buffer, or failure to apply the correct volume of buffer. For the oral fluid test, the most common errors were incorrect sampling technique during swabbing of the gum, and not transferring the device into the buffer solution.

Biosure and Chembio had the most spoiled tests. The Chembio and Biosure products use identical kit components and follow the same principle of testing; however, the kit components are packaged differently and have a different IFU design to align with Chembio and Biosure branding. The most common error seen across both products was related to the step: 'Push hard through the foil cap until fully seated in the buffer cap.' Those that made errors with this step had not pushed hard through the foil cap, and only inserted the tip of the device into the buffer which resulted in an inactive test and invalid result (no lines on test strip).

Performance assessment

Only participants who successfully achieved a self-test result on their own (3367/3600 (93.5%); range: 816/900 (90.7%) to 877/900 (98.2%)) were included in the performance calculation for clinical sensitivity and specificity; any incomplete tests or quits were not used to calculate the device performance. In total, there were 498 (14.8%) true positive HIVSTs (positive for both HIVST and ELISA), 7 (0.2%) false positive HIVSTs (positive for HIVST, negative for ELISA) 2853 (84.7%) true negative HIVSTs (negative for both HIVST and ELISA) and 9 (0.3%) false negatives (negative for HIVST, positive for ELISA). This resulted in an average sensitivity of 98.2% and a specificity of 99.8%, while also diagnosing 507 (15.1%) HIV-positive (sum of the true positives and false negatives) participants from the general population. The individual HIVST results are presented in Table 2.

Discussion

While previous studies have evaluated the usability of HIVSTs with contrived results, this report is the first South African report on the clinical performance of multiple devices with real-time results interpretation. The results of this study add to the growing body of evidence that supports the use of HIVSTs as a user-friendly and accurate testing approach to reach populations that may not have access to traditional clinic-based testing. A 2018 systematic review assessed the reliability of HIVSTs from 20 reports across 16 studies conducted between 1995 and 2016. In this review, 16 (80%) had a specificity greater than 98%, and although sensitivity varied substantially, 18 (90%) of the reports had a sensitivity greater than 80%.[16] Furthermore, an Orasure study from Singapore in 2012 (n=994) achieved a similar sensitivity of 97.4% and a specificity of 99.9%.[17] Another recent study of INSTI in Kenya (n=354) also revealed comparable results to our study with a sensitivity of 98.99% and a specificity of 98.15%.[18] A total of 330 (94.29%) participants found the device was easy to use, and the 15.1% of participants who tested positive in this study was slightly higher than the national prevalence of 13.1%.[19]

While corroborating previous results[15-18], this South African study demonstrates the sensitivity and specificity values of four HIVSTs to be higher than those attained during performance measurement for FDA approval[17], with a substantial sample size as outlined in the requirement for WHO prequalification. The National Department of Health in South Africa requires that any HIVST it procures or that is used on their sites must be approved by the South African Health Products Regulatory Authority (SAHPRA) or be prequalified by the WHO.

Table 1: Participant demographics

Demographic	Biosure	Orasure	INSTI	Chembio	Total
	Frequency (%)	Frequency (%)	Frequency (%)	Frequency (%)	Frequency (%)
Sample size	**900 (100.0)**	**900 (100.0)**	**900 (100.0)**	**900 (100.0)**	**3600 (100.0)**
Age					
18–25 years old	418 (46.4)	339 (37.7)	501 (55.7)	425 (47.2)	1683 (46.8)
26–35 years old	292 (32.4)	326 (36.2)	255 (28.3)	286 (31.8)	1159 (32.2)
Over 35 years old	190 (21.2)	235 (26.1)	144 (16.0)	189 (21.0)	758 (21.1)
Gender					
Female	419 (46.6)	383 (42.6)	460 (51.1)	394 (43.8)	1656 (46.0)
Male	481 (53.4)	517 (57.4)	440 (48.9)	506 (56.2)	1944 (54.0)
Nationality					
South African	820 (91.1)	745 (82.8)	829 (92.1)	807 (89.7)	3201 (88.9)
Zimbabwean	76 (8.5)	117 (13.0)	52 (5.8)	78 (8.7)	323 (9.0)
Other	4 (0.4)	38 (4.2)	19 (2.1)	15 (1.6)	76 (2.1)
Education Level					
Primary school or less	30 (3.3)	35 (3.9)	18 (2.0)	33 (3.7)	116 (3.2)
Secondary school	543 (60.3)	561 (62.3)	404 (44.9)	548 (60.9)	2056 (57.1)
Tertiary school (any)	327 (36.4)	304 (33.8)	478 (53.1)	319 (35.4)	1428 (39.7)
Employment Status					
Employed	211 (23.4)	208 (23.1)	149 (16.6)	285 (31.7)	853 (23.7)
Unemployed	581 (64.6)	618 (68.7)	647 (71.9)	433 (48.1)	2279 (63.3)
Student	107 (11.9)	74 (8.2)	104 (11.5)	182 (20.2)	467 (13.0)

Table 2: HIV self-testing (HIVST) usability and performance outcomes

Usability	Biosure (n=900)	Orasure (n=900)	INSTI (n=900)	Chembio (n=900)	Total (n=3600)
	Frequency (%)	Frequency (%)	Frequency (%)	Frequency (%)	Frequency (%)
Spoiled tests					
Invalid device	0 (0)	0 (0)	3 (0.3)	1 (0.1)	4 (0.1)
Required assistance	0 (0)	7 (0.8)	0 (0)	0 (0.0)	7 (0.2)
Quit	6 (0.7)	3 (0.3)	3 (0.3)	0 (0.0)	12 (0.3)
Collection error	36 (4.0)	7 (0.8)	31 (3.4)	27 (3.0)	101 (2.8)
Process error	60 (6.7)	11 (1.2)	15 (1.7)	74 (8.2)	160 (4.4)
Total	84 (9.3)	23 (2.6)	51 (5.7)	75 (8.3)	233 (6.5)
Successful HIVSTs	**816 (90.7)**	**877 (98.2)**	**849 (94.3)**	**825 (91.7)**	**3367 (93.5)**
Performance	**Biosure (n=816)**	**Orasure (n=877)**	**INSTI (n=849)**	**Chembio (n=825)**	**Total (n=3367)**
	Frequency (%)	**Frequency (%)**	**Frequency (%)**	**Frequency (%)**	**Frequency (%)**
True positive	126 (15.4)	152 (18.6)	98 (11.5)	122 (14.8)	498 (14.8)
True negative	687 (84.2)	717 (87.9)	750 (88.3)	699 (84.7)	2853 (84.7)
False positive	0 (0.0)	7 (0.9)	0 (0.0)	0 (0.0)	7 (0.2)
False negative	3 (0.4)	1 (0.1)	1[a] (0.1)	4 (0.5)	9 (0.3)
Outcomes	**Biosure (n=816)**	**Orasure (n=877)**	**INSTI (n=849)**	**Chembio (n=825)**	**Total (n=3367)**
	(%)	**(%)**	**(%)**	**(%)**	**(%)**
Usability index	95.9	97.4	97.1	97.8	97.1[b]
HIVST sensitivity	97.7	99.3	99.0	96.8	98.2[c]
HIVST specificity	100.0	99.4	100.0	100.0	99.8[c]

[a]One indeterminate ELISA result excluded, unable to recall participant for re-testing. Participant was conditionally diagnosed as HIV negative, as all three rapid tests (HIVST and both professional tests were negative).

[b]Usability was product specific, so direct comparisons between products should not be inferred.

[c]Total sensitivity and selectivity calculation with total TP, TN, FP and FN, not averages.

The high sensitivity and specificity of each HIVST evaluated in this study suggests that they should all be considered for approval, as they also meet all of the other standards outlined by the WHO prequalification documents. Each batch of devices was manufactured under ISO 14385 standards required for the design and manufacture of medical devices and each HIVST included IFUs with minimal language and simple pictorial instructions. At the time of this publication submission, two of the four devices in this assessment, OraSure and INSTI, had been prequalified by the WHO using data generated in this study.[20] Subsequent to this study Chembio also received prequalification. Data from these studies have been separately shared with SAHPRA, the South African National Department of Health and the South African National Institute for Communicable Diseases in order to facilitate the approval and usage of the products in implementation programmes such as the Self-Test Africa (STAR) project.

Despite the high levels of sensitivity and specificity, there were a number of user errors (notably with Biosure and Chembio), highlighting areas for improvement. Refining and tailoring the IFU to target markets (an action consequently implemented by Biosure and Chembio) and simplifying the device design could increase the overall usability of the device, thus further minimising errors. Whilst errors are expected in the hands of untrained users, it is imperative that users are able to recognise that an error has been made, and that the test invalidates itself, i.e. no control line/dot appears when a critical error is made. Tests which do not have specimen control lines, and produce control lines in the absence of any human specimen, can prove to be detrimental to HIVST as it could lead to an increase in false negative results. In order to build from these results and create a more robust body of evidence, future testing should be conducted with, and opinions elicited from, more diverse groups that include wider demographics and participants who are recruited independently of a clinical setting.

Limitations

This study has several limitations. A selection bias may have been created with convenience sampling, and while the evaluation of the devices in series ensured no cross-contamination, the general population may have become more aware of HIVST by the time the last device was tested, due to limited but expanding media coverage. The readability and comprehension of test instructions (we used only English IFUs for this evaluation) may be context and population specific, which limits the generalisation of these findings. Furthermore, an observation bias may be present, as the study was conducted under observation in a clinical setting, instead of alone in their homes.

Similar to the limitations of the HSTAR001 usability study, there is no validated or standardised usability test for HIVSTs, so the product-specific semi-structured questionnaire from HSTAR001 was used to quantify usability.[15] No direct comparisons could be made because of the different device components and non-standardised IFUs across kits. The sensitivity and specificity of each test also do not allow for direct comparisons, as these results were independently benchmarked against a gold standard, and not each other.

A fifth HIVST, Atomo, withdrew from the study halfway through data collection, so these results were not included in the aggregated data, or explored in the discussion, but the manufacturer did independently receive WHO prequalification for the device after withdrawing from the study.[21]

Conclusions

The four devices that were fully evaluated in this study and performed well, are among a growing number of HIVSTs intended to enter the South African market; OraSure, Chembio and INSTI have already received their WHO prequalification[22] and Biosure also received approval for use in South Africa. The results of this HSTAR003 performance evaluation methodology may also be used to guide similar evaluations among different populations. In the coming years, various HIVSTs will gain approval and enter the marketplace, which means that policies and distribution channels must be appropriately developed to accommodate this influx.

Acknowledgements

The HSTAR team would like to acknowledge Tanya Shewchuk and BMGF for funding; Mickey Urdea and Halteres Associates for study support; Rachel Baggaley and Cheryl Johnson for WHO HIVST technical support and Anita Sands and Robyn Meurant for engagement on assessment design and development of the WHO prequalification HIVST TSS.

Competing interests

Manufacturers provided input into study design; however, all testing and analysis was done independently of device manufacturers. W.D.F.V. and M.M. are both members of the WHO HIVST Technical Working Group and South Africa ST guidelines committee. All other authors have no conflicts of interest to declare.

Authors' contributions

M.M.: Conceptualisation, methodology, data collection, sample analysis, writing – revisions, project leadership, project management, funding acquisition. W.D.F.V.: Conceptualisation, methodology, writing – revisions, project leadership, funding acquisition. N.R.: Methodology, data collection, sample analysis, data analysis, validation, data curation, writing – revisions, project management. V.M.: Data collection, sample analysis, data analysis, validation, data curation, writing – revisions. S.T.L.E.: Data analysis, validation, data curation, writing – the initial draft, writing – revisions. A.E.F.: Data analysis, validation, data curation, writing – the initial draft, writing – revisions. L.M.: Data analysis, validation, data curation, writing – the initial draft, writing – revisions.

References

1. UNAIDS. Global AIDS update [webpage on the Internet]. c2017 [cited 2019 Sep 08]. Available from: http://www.unaids.org/en/resources/documents/2017/2017_data_book
2. Human Sciences Research Council. The Fifth South African National HIV Prevalence, Incidence, Behaviour and Communication Survey, 2017: HIV impact assessment summary report. Cape Town: HSRC Press; 2018. Available from: http://www.hsrc.ac.za/uploads/pageContent/9234/SABSSMV_Impact_Assessment_Summary_ZA_ADS_cleared_PDFA4.pdf
3. Johnson C, Kennedy C, Fonner V, Siegfried N, Figueroa C, Dalal S, et al. Examining the effects of HIV self-testing compared to standard HIV testing services: A systematic review and meta- analysis. J Int AIDS Soc. 2017;20(1), Art. #21594. https://doi.org/10.7448/IAS.20.1.21594
4. Wong V, Johnson C, Cowan E, Rosenthal M, Peeling R, Miralles ME, et al. HIV self-testing in resource-limited settings: Regulatory and policy considerations. AIDS Behav. 2014;18(Suppl 4):S415. https://doi.org/10.1007/s10461-014-0825-9
5. Kalibala S, Tun W, Cherutich P, Nganga A, Oweya E, Oluoch P. Factors associated with acceptability of HIV self-testing among health care workers in Kenya. AIDS Behav. 2014;18(Suppl 4):S405–S414. https://doi.org/10.1007/s10461-014-0830-z
6. Pai NP, Behlim T, Abrahams L, Vadnais C, Shivkumar S, Pillay S, et al. Will an unsupervised self-testing strategy for HIV work in health care workers of South Africa? A cross sectional pilot feasibility study. PLoS ONE. 2013;8(11), e79772. https://doi.org/10.1371/journal.pone.0079772
7. Pai NP, Sharma J, Shivkumar S, Pillay S, Vadnais C, Joseph L, et al. Supervised and unsupervised self-testing for HIV in high-and low-risk populations: A systematic review. PLoS Med. 2013;10, e1001414. https://doi.org/10.1371/journal.pmed.1001414
8. Richter M, Venter WD, Gray A. Enabling HIV self-testing in South Africa. S Afr J HIV Med. 2016;13(4):186–187. https://doi.org/10.7196/sajhivmed.858
9. Indravudh P, Choko A, Corbett E. Scaling up HIV self-testing in sub-Saharan Africa: A review of technology, policy and evidence. Curr Opin Infect Dis. 2018;31(1):14–24. https://doi.org/10.1097/QCO.0000000000000426
10. Choko AT, MacPherson P, Webb EL, Willey BA, Feasy H, Sambakunsi R, et al. Uptake, accuracy, safety, and linkage into care over two years of promoting annual self-testing for HIV in Blantyre, Malawi: A community-based prospective study. PLoS Med. 2015;12, e1001873. https://doi.org/10.1371/journal.pmed.1001873

11. Johnson C, Baggaley R, Forsythe S, Van Rooyen H, Ford N, Mavedzenge SN, et al. Realizing the potential for HIV self-testing. AIDS Behav. 2014;18(Suppl 4):S391–S395. https://doi.org/10.1007/s10461-014-0832-x

12. World Health Organization (WHO). Guidelines on HIV self-testing and partner notification: Supplement to consolidated guidelines on HIV testing services. Geneva: WHO Press; 2016.

13. South African National Department of Health. National HIV self screening guidelines 2018. Pretoria: Department of Health; 2018.

14. World Health Organization (WHO). Technical specifications series for submission to who prequalification – Diagnostic assessment: Human immunodeficiency virus (HIV) rapid diagnostic tests for professional use and/or self- testing. Geneva: WHO Press; 2017.

15. Majam M, Mazzola L, Rhagnath N, Lalla-Edward ST, Mahomed R, Venter WDF, et al. Usability assessment of seven HIV self-test devices conducted with lay-users in Johannesburg, South Africa. PLoS ONE. 2020;15(1), e0227198. https://doi.org/10.1371/journal.pone.0227198

16. Figueroa C, Johnson C, Ford N, Sands A, Dalal S, Meurant R, et al. Reliability of HIV rapid diagnostic tests for self-testing compared with testing by health-care workers: A systematic review and meta-analysis. Lancet HIV. 2018;5(6):e277–e290. http://dx.doi.org/10.1016/S2352-3018(18)30044-4

17. Tek Ng O, Chow A, Lee V, Chen MIC, Win MK, Tan HH, et al. Accuracy and user-acceptability of HIV self-testing using an oral fluid-based HIV rapid test. PLoS ONE. 2012;7(9), e45168. https://doi.org/10.1371/journal.pone.0045168

18. Bwana P, Ochieng L, Mwau M. Performance and usability evaluation of the INSTI HIV self- test in Kenya for qualitative detection of antibodies to HIV. PLoS ONE. 2018;13(9), e0202491. https://doi.org/10.1371/journal.pone.0202491

19. Statistics South Africa. Mid-year population estimates 2018. Statistical release P0309.3, 2018.

20. Ingold H, Mwerinde O, Ross AL, Leach R, Corbett E, Hatzold K, et al. The Self-Testing AfRica (STAR) Initiative: Accelerating global access and scale-up of HIV self-testing. J Int AIDS Soc. 2019;22(Suppl 1), e25249. https://doi.org10.1002/jia2.25249

21. World Health Organization (WHO). WHO Prequalification of in vitro diagnostics PUBLIC REPORT. Product: Mylan HIV Self Test (PQDx0320-090-00). Geneva: WHO; 2019. Available from: https://www.who.int/diagnostics_laboratory/evaluations/pq-list/191003_amended_pqpr_0320_090_00_mylan_hiv_self_test_v2.pdf?ua=1

22. World Health Organization (WHO). WHO Prequalification of in vitro diagnostics programme PUBLIC REPORT. Product: SURE CHECK HIV Self-Test (PQDx0054-006-01). Geneva: WHO; 2017. Available from: https://www.who.int/diagnostics_laboratory/evaluations/pq-list/191129_pqdx_0054_006_01_sure_check_hiv_self_test.pdf?ua=1

Hominin lower limb bones from Sterkfontein Caves, South Africa (1998–2003 excavations)

We describe late Pliocene and early Pleistocene hominin fossils from Sterkfontein Caves (South Africa), including two femoral specimens, as well as a partial tibia and a partial fibula. The fossils are likely assignable to *Australopithecus africanus* and/or *Australopithecus prometheus* and the morphology of each corroborates previous interpretations of Sterkfontein hominins as at least facultative bipeds.

Significance:

- A recent series of papers by our research team describes the morphology of a hominin skeleton from Sterkfontein Caves (South Africa), nicknamed 'Little Foot'. Based on its unique skull morphology, R.J. Clarke, the skeleton's discoverer, places it in the species *Australopithecus prometheus*, as distinct from the better-known and co-occurring *Australopithecus africanus*. Here we describe additional hominin thigh and leg fossils from Sterkfontein that, when considered in a comparative context, support the hypothesis that there was significant (probably interspecific) variation in South African hominin postcranial morphology during the late Pliocene and early Pleistocene.

AUTHORS:
Travis Rayne Pickering[1,2]
Jason L. Heaton[3]
Ron J. Clarke[2]
Dominic Stratford[4]
A.J. Heile[1]

AFFILIATIONS:
[1]Department of Anthropology, University of Wisconsin-Madison, Madison, Wisconsin, USA
[2]Evolutionary Studies Institute, University of the Witwatersrand, Johannesburg, South Africa
[3]Department of Biology, Birmingham-Southern College, Birmingham, Alabama, USA
[4]School of Geography, Archaeology and Environmental Studies, University of the Witwatersrand, Johannesburg, South Africa

CORRESPONDENCE TO:
Travis Pickering

EMAIL:
travis.pickering@wisc.edu

DATES:
Received: 26 July 2019
Revised: 01 July 2020
Accepted: 02 July 2020
Published: 29 Jan. 2021

HOW TO CITE:
Pickering TR, Heaton JL, Clarke RJ, Stratford D, Heile AJ. Hominin lower limb bones from Sterkfontein Caves, South Africa (1998–2003 excavations). S Afr J Sci. 2021;117(1/2), Art. #6758. https://doi.org/10.17159/sajs.2021/6758

ARTICLE INCLUDES:
☒ Peer review
☐ Supplementary material

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

EDITOR:
Margaret Avery

KEYWORDS:
Australopithecus, femur, tibia, fibula

FUNDING:
University of Wisconsin-Madison, South African National Research Foundation

Introduction

A long history of palaeoanthropological research at Sterkfontein Caves (Gauteng, South Africa), starting in 1936, has yielded a massive store of hominin fossils.[1-5] As one in a series of papers, we report here on four previously undescribed hominin fossils excavated by R.J.C. between 1998 and 2003. Those other papers present descriptions and interpretations of sizeable samples of teeth[6] and of axial and upper limb bones[7,8]. The sample described here comprises fewer materials, representing two femur specimens, a partial tibia and a partial fibula. All four specimens preserve sufficient anatomy to indicate that the individuals from which they derived were bipeds, but more comprehensive functional interpretations of the fossils are limited by their fragmentary nature. Each also preserves taphonomic information that we report.

Two of the fossils derive from Member 4 of the Sterkfontein Formation, the site's most hominin-rich deposit, which is probably slightly greater than 2.6 to around 2.5 million years old (Ma) (Figure 1).[9-13] The other two are part of the Jacovec Cavern fossil assemblage, which includes a modest sample of hominin specimens, previously described by Clarke[14]. The original interpretation of the Jacovec stratigraphy identified several generations of cave fill, including three sedimentary units identified as 'orange', 'brown', and 'stony'.[14] More recent research demonstrates multigenerational infilling and reworking of the chamber's sediments, which created several secondary deposits and a talus composed of orange and brown units.[15] Dating of the Jacovec deposits is still in progress. Based on the presence of a fossil attributed to *Equus*, Kibii[16] proposed a relatively young age of 2.34 Ma for the recovered faunal assemblage. However, Partridge et al.[14] provided an age of 4.02±0.27 Ma for orange sediments exposed on the chamber ceiling, which yielded a partial hominin cranium cataloged as StW 578. It is concluded that all hominin fossils thus far recovered from the chamber floor derive from this in-situ orange unit.[14,15] Accordingly, we propose that until such time as multiple, stratigraphically associated, and dateable proxies contradict this radiometric date, it remains the best plausible estimate for the age of the Jacovec fossils. Based on craniodental analyses, Clarke[5,17-19] has argued for the presence of two coeval hominin species in Member 4: *Australopithecus africanus* and *Australopithecus prometheus*. Additionally, analysis of hominin teeth from Jacovec indicate that both taxa also occur in that deposit.

Methods

We provide information on each fossil's spatial context, its relative completeness and condition, and its morphology and osteometrics. The legend to Figure 1 summarises the excavation coordinate system employed at Sterkfontein. As to condition, we recorded the degree to which a specimen suffered subaerial weathering, using Behrensmeyer's[20] well-known weathering stage system, and noted its degree of staining by manganese dioxide. Bone surface modifications were also identified using 10x power magnification.[21] All fractured bone surfaces were assessed with reference to the 'angle formed by the fracture surface and bone cortical surface'[22 (p.34)]. Typically, fracture angles on long limb bones that were created when the bone was 'green' (i.e. before significant loss of a bone's organic fraction and its desiccation) are usually either acute or obtuse, while those created on dry long limb bones are usually right angles.[22-24] We used Mitutoyo™ digital calipers to collect standard osteometric linear measurements and collected standard osteometric angular measurements with SPI™ 0–180° protractors.[25-27] We followed recently published methods[28] in order to derive femoral neck anteversion on the single proximal femur specimen in the analysed sample.

Results

StW 598 (Jacovec Cavern)

This proximal left femur was described preliminarily by Clarke[14]. It is an exquisite specimen, stained uniformly by manganese dioxide but essentially unweathered (stage 0) and preserving the head and nearly half of the diaphysis (maximum length = ~153.0 mm) (Figure 2). The fossil is truncated distally along a relatively straight, right-angled breakage plane, indicating that the bone was leached of much, if not all, of its organic content when it was fractured.

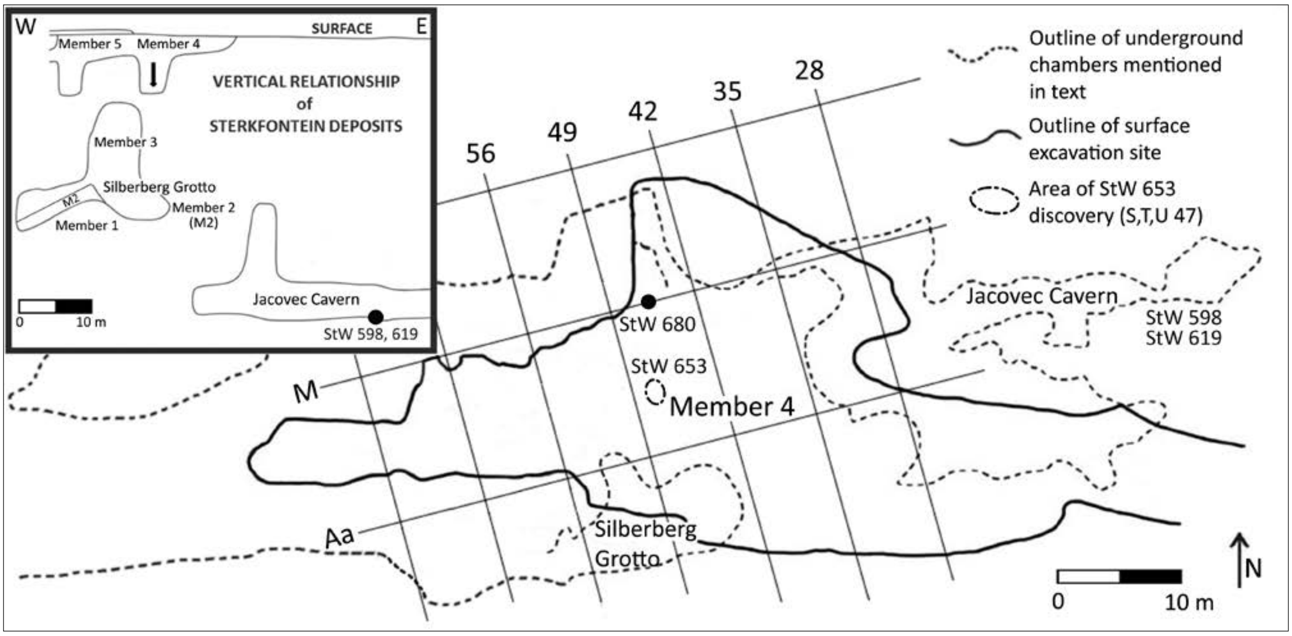


Figure 1: Composite plan view of the eastern portion of the Sterkfontein fossil site, with surficial deposits indicated by solid lines and underground deposits delimited by dashed lines (main image) (modified from Reynolds and Kibii[64] Figure 2 and references therein), and a schematic ~EW profile illustrating the vertical relationships of some of the site's major deposits (inset) (modified from Clarke[65] Figure 6). All fossils described in this paper derive from Member 4 and Jacovec Cavern. Sterkfontein sediments are excavated in spits of 3' x 3' x 1' volume under an alpha (NS coordinate)-numerically (EW coordinate) labelled grid; values in feet and inches listed in fossil specimen descriptions indicate depths below site datum.[66]

Further, the fracture surface is partially coated by breccia, attesting to the antiquity of the breakage event that created it. A crack emanates from the middle of the anterior edge of this fracture surface and travels proximally, along the middle of the anterior surface of the diaphysis, for a length of ~80.0 mm. The crack is wider mediolaterally for its distal ~two-thirds than it is proximally. A similarly constructed longitudinal crack courses the middle of the posterior diaphysis (along the medial lip of the linea aspera), starting distally in the middle of the dorsal edge of the specimen's distal fracture surface and terminating ~78.0 mm proximal to that point, near the junction of the base of the lesser trochanter and the proximal terminus of the pectineal line. Both cracks are infilled by breccia, which caused their mediolateral expansions, especially distally. Other ancient modification to StW 598 occurs at its proximolateral corner. Unlike the previously described incidences of damage, this alteration of the fossil was most likely inflicted when the bone was still fresh and its trabeculae were packed with nutritionally attractive red marrow and grease. The crenulated surface[3] of the base of the largely missing greater trochanter, as well as several deeply invasive tooth gouges in the exposed trabeculae at the anterolateral portion of the distal margin of that damaged surface, indicate that the trochanter was chewed away by a prehistoric carnivore(s). This carnivore-inflicted damage continues superomedially across the anteroproximal femoral neck, where several layers of lamellae were peeled away in a dorsoventrally wide strip (maximum width = ~14.0 mm) that terminates ~11.0 mm from the superior rim of the femoral head. The head of StW 598 is small and spherical, with a relatively large fovea capitis that is placed dorsally on the medial aspect of the joint surface (Table 1). The fovea capitis is deep, with a superior and dorsal margin that is contiguously sharply raised. The femoral head is hafted to the bone's diaphysis at a neck-shaft angle of 121°, via a mediolaterally elongated and anteroposteriorly compressed neck (Table 1). StW 598 has a fairly marked femoral anteversion angle of 102° (Figure 3). The obturator externus groove of the specimen is not palpable superiorly and just barely so dorsolaterally. Likewise, the specimen's intertrochanteric crest and intertrochanteric and spiral lines are underdeveloped. In contrast, the fossil shows a small but robust, superoinferiorly elongated (maximum superoinferior length = ~16.3 mm; maximum mediolateral width = ~10.4 mm) lesser trochanter, a well-developed pectineal line and an especially rugose gluteal tuberosity. The lesser trochanter projects slightly medially beyond the medial margin of the proximal diaphysis, so that it (the tubercle) is just visible when viewing the anterior aspect of the bone. The gluteal marking is a well-developed, posterolaterally placed ridge that merges with the pectineal line ~49.0 mm distal to the inferior margin of the lesser trochanter, at a small, distally opening nutrient foramen, to form the superior terminus of the mediolaterally broad linea aspera. Viewed medially or laterally, the subtrochanteric diaphysis appears relatively flat anteriorly (especially proximally) and only slightly more convex posteriorly, forming a gentle posterolaterally placed angle along the course of the linea aspera (Table 1). The shaft cross-section is roughly circular at the distal fracture edge, with an anteroposterior diameter of 20.5 mm and mediolateral diameter of 21.1 mm at that level (these dimensions are uncorrected for the slight mediolateral gaps in the antero- and posterodistal diaphysis described above). The thickest portion of the cortex at this natural fracture surface is posteriorly, at the position of the linea aspera (7.4 mm). Because the diaphyseal midpoint cannot be determined precisely on the broken specimen, we cannot report a pilasteric index (midshaft anteroposterior diameter divided by midshaft mediolateral diameter, multiplied by 100[29]). However, an index of 97 is calculated using these measurements taken from the distalmost edge of the fossil.

StW 619 (Jacovec Cavern)

This partial distal epiphysis of a left femur is heavily mottled by deposits of manganese dioxide but shows little subaerial weathering (stage 0) (Figure 4). All areas of damage on the fossil are ancient, as evidenced by the fact that the exposed trabeculae are smoothed and polished and are filled by red cave sediments. The specimen is broken proximally, along a relatively straight transverse plane, at the inferior margin of the distal metaphysis. The edge of this fracture surface is roughly right-angled for its entire circumference; two short but deep, probable carnivore, tooth scores emanate in an inferoanterior direction from the medial margin of this fracture edge. Most of the lateral condyle and lateral epicondyle of StW 619 are missing, although much of a roughened impression for the lateral head of the gastrocnemius is preserved. The lateral patellar lip of the specimen is broken away laterally and proximally but it is obvious from an inferior view that the lateral lip projected significantly anteriorly (Figure 4). The region of the lateral meniscal groove is missing.

Table 1: Metric comparisons of South African early hominin proximal femora[a]

Specimen/ Taxon	Head SI (mm)	Neck SI (mm)	Neck AP (mm)	Neck shape[b]	Neck length (mm)	Subtrochanteric diaphysis AP (subAP) (mm)	Subtrochanteric diaphysis ML (subML) (mm)	Meric Index[c]
MLD 46[d]	[38.4]	30.1	19.8	65.8	36.5	–	–	–
SK 82[e]	34.4	27.2	19.2	70.6	41.7	24.8	30.4	81.6
SK 97[e]	37.1	26.2	19.6	74.8	46.8	23.3	35.3	66.0
SK 3121[e]	28.6	19.1	16.0	83.8	–	–	–	–
SKW 19[e]	30.2	–	–	–	–	–	–	–
SWT1/LB-2[e]	34.4	25.9	17.7	68.3	35.3	–	–	–
Sts 14[d]	–	22.6	13.4	59.3	–	–	–	–
StW 25[d]	[31.3]	[22.9]	–	–	–	–	–	–
StW 99[d]	[34.6]	29.6	18.5	62.5	[49.9]	–	–	–
StW 311[d]	35.8	26.4	20.5	77.7	–	–	–	–
StW 392[d]	31.4	[21.4]	–	–	–	–	–	–
StW 403[d]	–	24.9	17.3	69.5	–	–	–	–
StW 479[d]	–	[23.5]	17.7	75.3	–	–	–	–
StW 501[d]	[31.8]	[23.8]	[16.5]	69.3	–	–	–	–
StW 522[d]	30.8	20.9	15.5	74.2	31.3	–	–	–
StW 527[d]	[33.3]	–	–	–	–	–	–	–
StW 598	32.1	25.5	16.5[f]	64.7	40.4	20.2	22.3	90.6

[a]*Standard linear measurements from McHenry and Corruccini[27]. Abbreviations: SI = superoinferior; AP = anteroposterior (all SIs and APs are diameters); MLD = Makapansgat Limeworks Dump (South Africa); SK, SKW and SWT1/LB = Swartkrans (South Africa); Sts and StW = Sterkfontein. Measurements in brackets are estimates; dash in a cell indicates that the measurement was unobtainable.*

[b]*Neck shape = neck AP divided by neck SI, multiplied by 100.*

[c]*Meric index = subAP divided by subML, multiplied by 100.*

[d]*Head SI, neck SI, neck AP and neck length values from Harmon[46].*

[e]*Head SI, neck SI, and neck AP values from Pickering et al.[67]; subAP and subML values from Robinson[30].*

[f]*Uncorrected for missing cortical bone on anteroproximal aspect of neck.*

Figure 2: The StW 598 hominin left proximal femur from Jacovec Cavern, Sterkfontein, shown in, from left to right, superior view (anterior facing up) and anterior, medial, posterior, and lateral views (superior facing up in all views). Bar scale = 1 cm.

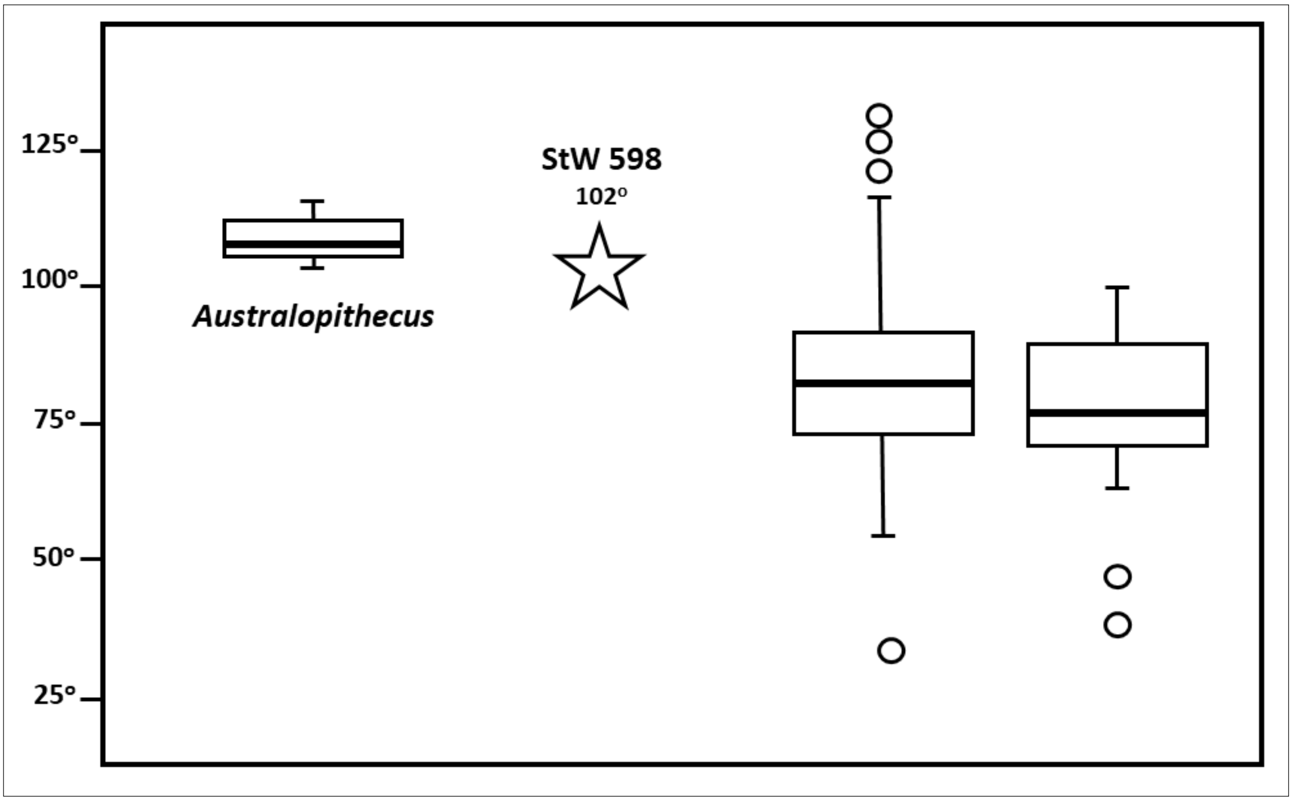


Figure 3: Head/neck anteversion of the StW 598 proximal femur from Jacovec Cavern, Sterkfontein, compared to that of other *Australopithecus* and extant hominoid femora. Modified from Marchi et al.[28] Figure 23. The box-and-whisker plot shows the median (dark horizontal line), upper and lower quartiles (boxes), range (whiskers), and outliers (circles). Comparative data from Marchi et al.[28]: *Australopithecus* specimens include A.L. 288-1 and A.L. 333-95 (Hadar, Ethiopia), StW 99 (Sterkfontein, South Africa), and MH1 (Malapa, South Africa).

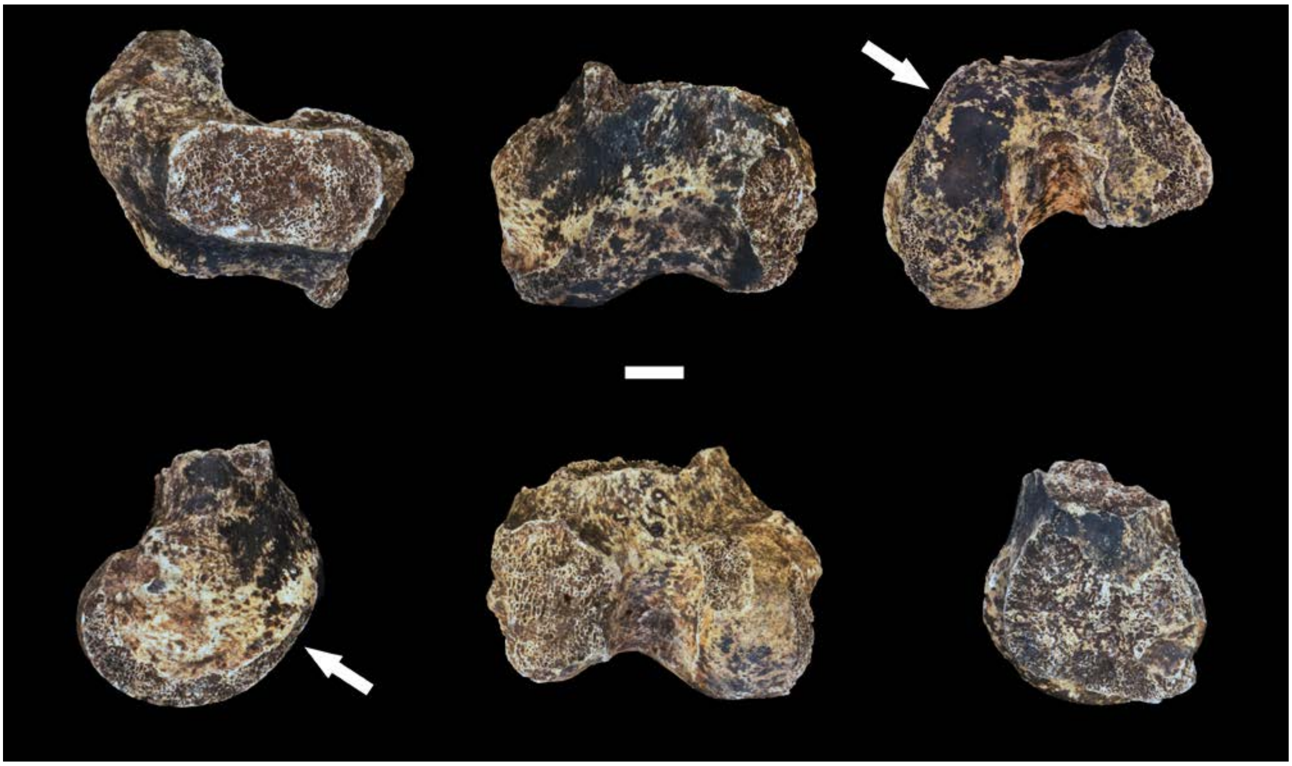


Figure 4: The StW 619 hominin left distal femur from Jacovec Cavern, Sterkfontein, shown in superior (posterior facing up), anterior (superior facing up), and inferior (anterior facing up) views (left to right, top row), and in medial, posterior and lateral views (superior facing up; left to right, bottom row) Arrows indicate the estimated position of the medial meniscal groove. Bar scale = 1 cm.

Several areas of the medial condyle are also eroded, including the entirety of its superior border from that feature's anterior point of connection with the medial patellar lip to its dorsal termination. This arc of damage extends laterally into the region in which the medial meniscal groove was positioned, obliterating the medial extent of that feature. We are, however, able to discern what we judge to be the lateralmost extremity of the groove as indicated by arrows in Figure 4. This probable groove remnant is quite shallow. StW 619 lacks a well-developed medial condylar boss. The distal medial condyle does not drop appreciably inferiorly dorsal to the presumed meniscal groove; instead, the distal medial condyle is fairly smooth across its extent, except for a low and mediolaterally narrow, anteroposteriorly elongated elevation along the length of its lateral edge. The adducter tubercle of StW 619 projects only modestly in superior direction, away from the solid mass it forms in connection medially with the blunt, strongly projecting medial epicondyle. In medial view, the medial condyle is elliptical in shape. Maximum (non-anatomical; taken mediolaterally) linear length of StW 619 is 55.5 mm. In general form and size, StW 619 is very similar to Sts 34 and TM 1513 – two hominin distal femora from Sterkfontein Member 4 that also preserve medial condyles (Table 2).[30]

Table 2: Metric comparisons of hominin distal femora (preserving medial condyles) from Sterkfontein[a,b]

	Femur specimens		
Standard linear measurement	**Sts 34**	**StW 619**	**TM 1513**
Anteroposterior diameter of the distal shaft	29.0	25.5	27.5
Anteroposterior diameter of the medial condyle	45.5[c]	[43.0]	[45.0]
Mediolateral (transverse) diameter of medial condyle	–	[20.0]	[22.1]
Condylar notch width	16.7[c]	[13.5]	13.4[d]

[a]Standard linear measurements from McHenry and Corruccini[27]. Sts and StW = Sterkfontein; TM = Kromdraai (South Africa).

[b]All measurements in mm; measurements in brackets are estimates; dash in a cell indicates that the measurement was unobtainable.

[c]Agrees with measurement in Robinson[30].

[d]Disagrees with measurement in Robinson[30] (12.4 mm), which was taken anteriorly (contra recommendation in McHenry and Corruccini[27]).

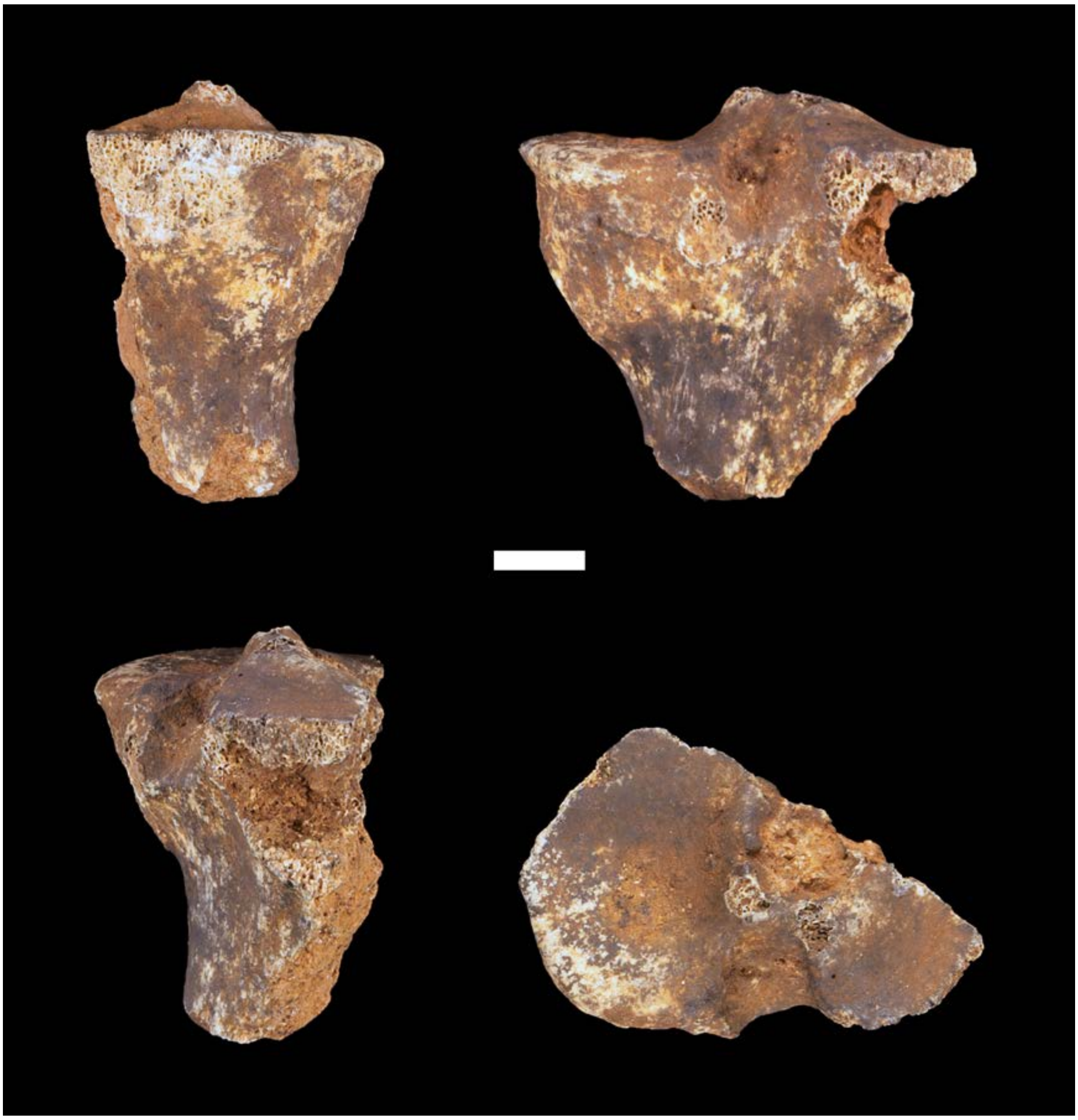


Figure 5: The StW 653 hominin right proximal tibia from Sterkfontein Member 4 shown in medial and posterior views (superior facing up in both; left to right, top row) and in lateral (superior facing up) and superior (anterior facing up) views (left to right, bottom row). Bar scale = 1 cm.

StW 653 (Member 4, mixed S,T,U/47)

This partial proximal epiphysis and extreme proximal metaphysis of a right tibia is well stained by manganese dioxide and is unweathered (stage 0) (Figure 5). Its maximum length is mediolateral at 49.7 mm. The anterolateral aspect of the specimen is missing and the exposed trabeculae in that area are filled with lightly calcified breccia. Dorsolaterally, the specimen's proximal end preserves damage reminiscent of ragged edge chewing imparted by carnivores (cf.Brain[3]), but there is no independent indication of feeding damage on the rest of the specimen in the form of tooth scores or pits. The distal fracture surface of StW 653 is obscured by adhering breccia but appears to be right-angled.

Most of the medial condyle of StW 653 is still intact, as is the approximate dorsal third of the lateral condyle. The medial condyle is dorsoventrally elongated (estimated anteroposterior diameter = 34.5 mm; estimated mediolateral diameter = 22.5 mm) and slightly concave, bounded by a blunt but superiorly projecting rim dorsomedially (the rest of the rim is missing so we cannot comment on its original form). The concavity of the medial condyle is accentuated by the acute manner in which the medial intercondylar tubercle rises from it superiorly. The intercondylar eminence and the posterior and most of the anterior intercondylar areas are also preserved, as is the lateral intercondylar tubercle. The last falls less severely distolaterally toward the lateral condyle than does the medial intercondylar eminence fall distomedially toward the medial condyle. The superior apices of both intercondylar eminences are chipped, exposing their underlying trabeculae. The small portion of lateral condyle is convex dorsally, appears moderately concave ventrally, and is positioned more superiorly than is the medial condyle; its dorsomedial corner appears to be marked by a small 'meniscal notch' (cf. Tardieu[31]). In sum, the morphology of the tibial plateau is reminiscent of those of 'typical' modern human tibiae. Distal to the plateau, the proximolateral branch of the soleal line shows a degree of development that is also comparable to those of modern human tibiae; the groove for the semimembranosus is humanlike in morphology (i.e. deep and circular ventrally; superoinferiorly shorter and anteroposteriorly elongated dorsally) and position, immediately at the base of the dorsomedial corner of the medial condyle. The attachment area for the medial collateral ligament is very rugged, projecting strongly medially from the proximal metaphysis, inferior to it.

StW 680 (Member 4, M/46 18'5'–19'5')

This partial diaphysis of a hominin fibula, measuring 84.3 mm in maximum length, is well stained by manganese dioxide but is unweathered (stage 0) (Figure 6). It lacks surficial tooth scores or pits, but one of its ends terminates in an irregular break reminiscent of carnivore-induced ragged-edge chewing. This end is densely packed with trabeculae, indicating that it is metaphyseal. This region lacks a diagonal curve of an anterior border or crest toward the position of a lateral malleolus, as would be predicted for a distal fibula portion. Accordingly, we conclude that this part of the specimen is from the proximal end of the bone. Superiorly, the cross-section of the fossil is roughly triangular, with two sides of the triangle being slightly convex and divided by a low ridge, while the third side is flat-to-concave, and bounded on each side by sharply defined crests. When viewed in standard anatomical position, the lateral and posterior sides of a typical hominin fibula are convex, while its medial side, lying between well-defined crests, is concave. Following this understanding, it is clear that StW 680 is a left fibula. In addition, the specimen compares favourably to two other hominin left fibulae from Sterkfontein, StW 356 (Member 4) and StW 573 (Member 2).

The distal termination of the specimen is spiral but its fracture edge is right-angled, indicating that the break was probably induced by static loading (e.g. sediment compaction) and when the specimen was at least partially degreased. Its diaphysis is more robust than those of the two other Sterkfontein fibula mentioned above, and its fibular neck is round in cross-section. Martin and Saller[26] recommend that diameters of the fibular neck be taken at that feature's smallest circumference. Following that guideline, the anteroposterior diameter of the neck of StW 680 is 11.0 mm and its mediolateral diameter is 10.5 mm. Distal to the neck, the medial and lateral surfaces diverge from each other at an acute angle formed by a low anterior border that is more elevated (i.e. anteriorly projecting) distally than it is proximally; the medial surface is flat proximally to very slightly concave distally for its preserved length, while the lateral surface bulges convexly for most of its length superiorly along a rounded ridge and is flat inferiorly. The posterior border, for its whole length, takes roughly the same rounded form as the lateral surface. In contrast, the interosseous border is sharp but still does not project all that markedly from the main body of the diaphysis.

Figure 6: The StW 680 hominin left fibula from Sterkfontein Member 4 shown in, left to right, posteromedial and lateral views (superior facing up). Bar scale = 1 cm.

Discussion

Neotaphonomic research, utilising modern baboon carcasses as proxies for early hominin cadavers, shows that the knee joint is especially susceptible to destruction by feeding carnivores.[32] In this context, it is worth mentioning that taphonomic studies have also concluded that carnivores played at least some role in the creation of the hominin fossil assemblages from Sterkfontein Member 4 and Jacovec.[4,7,14] We also note the following general characteristics of the South African hominin fossil record: proximal femora are more common than are distal specimens; proximal tibiae are rarer than are distal specimens; and, fibular specimens of any completeness are exceedingly exceptional.[2-5] In sum, it thus seems especially fortunate that the chewed (and possibly chewed) thigh and leg specimens described here survived to become part of the Sterkfontein fossil record.

Member 4 fossils

Bipedalism is a defining hominin characteristic and the knee obviously plays a central role in mammalian locomotion, so it is no surprise that palaeoanthropologists have paid particular attention to that joint. More specific to Sterkfontein, Berger and Tobias[33] claimed that the anteroposterior convexity of the lateral condyle of StW 514a, a hominin proximal tibia from Member 4, as well as the fossil's lack of a 'meniscal notch' might indicate that (at least one) *Australopithecus* at the site locomoted in a 'chimpanzee-like' manner. However, a study that compared digital, three-dimensional surface areas of the lateral tibial condyles of various extant African hominoids, South African *Australopithecu*s, and East African *Australopithecus afarensis* to two-dimensional surface area and arc and chord length measurements of the same feature concluded 'that tibial condylar curvature is a weak discriminator of locomotor variation in extant, and presumably fossil, hominoids'[34(p.124)]. Further, Tardieu[35] showed earlier that the lateral condyles of other early hominin tibiae – none of which would otherwise be characterised as 'chimpanzee-like' – also lack notches; more recently, the absence of lateral condylar notches has even been documented in some tibiae of fully bipedal modern humans.[36] Zipfel and Berger[37] acknowledged these observations in a study comparing StW 514a to another hominin proximal tibia from Sterkfontein Member 4, StW 396, but still emphasised, as did Berger and Tobias[33(p.343)] before them, that the semimembranosus attachment site of StW 514a 'forms a marked circular depression situated on the posteromedial margin of the condyle immediately below the medial condylar surface', implying that this enthesis morphology is distinctly chimpanzee-like. Our own observations of modern human tibiae disagree with this insinuation; in our experience, many modern human tibiae show large, strongly indented semimembranosus attachment sites with distinct borders, such as is illustrated in Figure 7. In sum, it thus seems that the StW 541a is not, in fact, particularly 'chimpanzee-like', but instead simply expresses a combined morphology that falls within the bounds of normal variation for hominin tibiae.

Figure 7: An approximate medial view of a modern human right proximal tibia showing a large, strongly indented and circular attachment site (circled), with a well-defined border, for the semimembranosus, contradicts the suggestion[33] that such morphology for that attachment site is uniquely 'chimpanzee-like.'

The same is true of StW 653, the new Member 4 hominin tibia described here, which compares quite favourably to not only other *Australopithecus* tibiae but also to those of modern humans. For example, the new fossil's strongly developed soleal line accords with those of fully bipedal modern humans. In no way does this suggest that StW 653 should be assigned to the genus *Homo*, but instead confirms the conclusions of others who question the efficacy of using certain morphological features to place early hominin tibia fossils into particular genera and/or species (see for example Dugan and Holliday[36]).

In contrast, the new StW 680 fibula from Member 4 differs markedly from those of modern humans and is instead more like those of extant African apes and also compares favourably to the Sterkfontein *Australopithecus* fibula fossils, StW 356 and StW 573. It possesses a generally triangular neck and has a high neck robusticity index (mediolateral diameter/anteroposterior diameter *100)[28] of 95.4, compared to a mean for two other *Australopithecus* proximal fibulae of 95.7 and a mean for 23 modern human fibulae of 81.8±16.6 (range = 52.6–111.5) (comparative data from Marchi et al.[28]). Moreover, the shape of the origin for the peroneus longus is convex, like those of other early hominin fibulae, including especially those attributed to *Australopithecus*, but unlike the origin on modern human fibulae.[28]

Jacovec fossils

As first noted by Clarke[14], in possessing a relatively small head and long neck, StW 598 resembles SK 82 and SK 97[30] – proximal femur fossils from Swartkrans (South Africa) that are usually assigned to *Paranthropus robustus*. StW 598 also resembles StW 99, a large femur specimen from Sterkfontein that is usually attributed to *Australopithecus*, but that – given its morphological continuity with the Swartkrans *Paranthropus* femur fossils and its possible origin from the *Paranthropus*-bearing Member 5 unit of the Sterkfontein Formation[13] – is likely actually also *Paranthropus* (Table 1). With its long neck, StW 598 is, however, very similar to definitive *Australopithecus* femur specimens, StW 479 and StW 367, from Sterkfontein Member 4. StW 598 and StW 367 are, in fact, so similar that upon their cursory comparison, one could understandably conclude that the fossils are antimeres. Further detailed observations prove that conclusion erroneous, but the salient point made is that the morphology of StW 598 is not unique in the hominin fossil record. Adding to the comparative complexity, StW 522, a presumptive *Australopithecus* femur from Sterkfontein, shows a small head and short neck (Table 1).[14]

StW 598 has a remarkably high meric index of 90.6 (Table 1), compared to a mean meric index of 74.8±3.5 (range = 66.4–81.7) for 18 South and East African *Australopithecus* and *Paranthropus* femora.[28] The cause(s) of femoral shaft shape remain hypothetical, but biomechanical explanations include reference to the influences of both the vasti and gluteal complexes. However, the meric index ranges of both knuckle-walking African apes (mean meric index of 42 *Pan troglodytes* femora = 85.0±5.5 [range = 71.1–95.5]; mean meric index of 47 *Gorilla gorilla* femora = 83.2±4.0 [range = 76.5–93.8][28]) and terrestrial bipedal modern humans (mean meric index of 195 *Homo sapiens* femora = 80.8±6.8 [range = 56.1–96.6][28]) not only overlap but also encompass the index value of StW 598. Thus, we are currently reluctant to extrapolate any functional interpretations of the fossil based on its round diaphyseal cross-section.

Likewise, although the position and narrow, ridge-like form of StW 598's gluteal enthesis suggest a relatively humanlike insertion of its gluteal musculature[38], without a preserved greater trochanter, it would be unwise to place too much explanatory emphasis on that muscle scar. Indeed, in many other features, StW 598 is quite dissimilar to the proximal femora of modern humans. For instance, its anteroposteriorly compressed, superoinferiorly tall, and strongly anteverted neck is unlike the femoral necks of extant *H. sapiens* and, instead, mirrors the morphology of nearly every other known *Australopithecus* femur specimen.[28,39]

Moving to the distal femur, the femora of bipedal hominins are/were adapted, via significant anterior projection (i.e. 'elevation') of the lateral patellar lip, to resist dislocation of the patella under the load of stance phase valgus. A simple way to visually assess the *functional* elevation of the lateral patellar lip of a femur is to view the specimen distally, with an axis running through the condylar meniscal grooves oriented horizontal to the flat base over which the femur is positioned.[40] Unfortunately, because of the fragmentary nature of the fossil, the meniscal axis of StW 619 cannot be estimated with precision. However, placing the medial patellar surface of the specimen horizontally in distal view 'corresponds closely to the orientation recommended by Lovejoy[40'41(p.1232999-1)] (Figure 8). This view clearly illustrates the appreciable depth of StW 619's patellar groove, as well as the significant functional elevation of its lateral patellar lip. Similarly, even though StW 619 lacks a discernible medial condylar boss, when the specimen is rotated into an estimated position of full extension and viewed distally, the anterior termination of the medial condyle assumes an ovoid shape. Lovejoy[40] emphasises that this shape, in this view, is an ancillary indication that the tibial contact area of a femoral condyle was elongated relative to those of the femora of quadrupedal mammals. Bipedal primates with such elongated, ellipsoid femoral condyles possess(ed) knees that are 'tibial dominant'.[40]

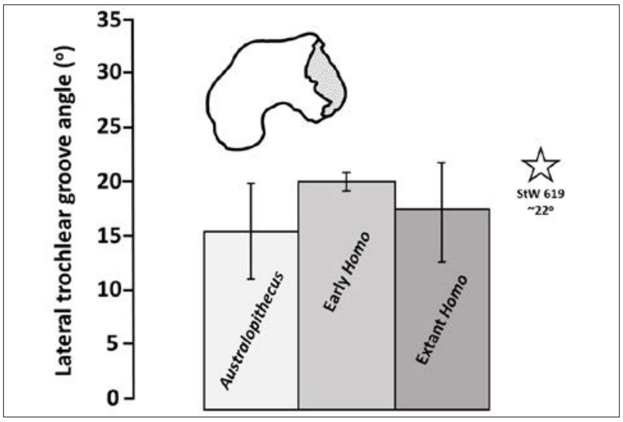


Figure 8: Comparison of the estimated lateral trochlear groove angle[68] of StW 619 to those of other hominin distal femora (modified from Desilva et al.[41] Figure 1). The distal view outline of StW 619 shown on the graph is positioned so that its medial patellar surface is horizontal to the straight bottom edge of the image, which approximates the view recommended[40] for determining the functional elevation of the lateral patellar lip (see discussion in text). Comparative specimens include: Sterkfontein *Australopithecus* (specimens Sts 34 and TM 1513) plus Hadar, Ethiopia, *Australopithecus* (specimens A.L. 129-1, A.L. 333-4 and A.L. 333w-56); Lake Turkana, Kenya, early *Homo* (specimens KMN-ER 1472, KNM-ER 1481, KNM-ER 15000) and extant *Homo sapiens* (data from Desilva et al.[41]).

The conclusion that the knee of the StW 619 hominin – although not an exact morphological match for a modern human knee – was nonetheless capable of full, modern humanlike bipedal extension, is not unexpected given a broader view of the hominin fossil sample from Jacovec. As discussed above, the proximal femur StW 598 shows derived features, such as an elongated neck, a low neck-shaft angle and a relatively small head, observed in other proven bipedal hominins.[38,41-48] In addition, Pickering et al.[7] (see also Partridge et al.[14]) have also described two lumbar vertebrae from Jacovec that indicate the hominin(s) from which they derived possessed intrinsic lumbar lordosis. Lumbar lordosis is, of course, a critical component of hominid bipedality[30], as it functions to maintain orthogrady by ameliorating strain on the dorsal spinal ligaments and by absorbing shock emanating from upright activities on terrestrial substrates[49-51].

None of this is to suggest that all of the Jacovec hominins were necessarily as fully committed to terrestrial bipedalism as are modern humans. The StW 605 manual proximal phalanx from Jacovec is too damaged to quantify its included angle but it appears fairly curved longitudinally.[8] Consensus posits a causal link between curved manual proximal phalanges and significant degrees of arboreal behaviour in primates.[52,53] Similarly, the StW 606 hominin clavicle from Jacovec shows a mix of modern humanlike and apelike features, the latter of which would have endowed the hominin from which it derived with good climbing abilities.[7,14]

Conclusion

When added to results from previous studies of *Australopithecus* postcranial samples from Sterkfontein[5,7,8,30,54-59], this study corroborates that those samples include functionally heterogeneous mixes of elements, some of which indicate postural and locomotor behaviours that are apelike and others of which indicate human-like adaptations. Starting from this basic understanding, there is, in our opinion, a fundamental problem that underlies many ensuing debates over the postcranial functional morphology of *Australopithecus*. Specifically, we are troubled by the fact that many disputants in these debates misunderstand or ignore the completely salient possibility (likelihood?) that the Sterkfontein Member 4 hominin postcranial collection samples at least two coeval species, *A. africanus* and *A. prometheus*, as is the case for the large craniodental sample of hominin fossils from that depositional unit.[5,6,17-19] For instance, Harmon[60] detected significant variation in the shape of proximal femora typically assigned to *A. africanus*, which she considered intraspecific variation, but that might actually be an interspecific difference. In contrast, Clarke[5] demonstrated that two hominin first metatarsals from Sterkfontein Member 4 show distinct morphologies indicative of different modes of locomotion and thus concluded that the fossils likely represent separate species (see also Deloison[61]). Additionally, Kibii and Clarke[57] suggested that the pelves of the partial Sterkfontein Member 4 skeletons, Sts 14 and StW 431, do not necessarily sample the same species. This issue has been clarified recently with the cleaning and reconstruction of the StW 573 skeleton of a female *A. prometheus* from Sterkfontein Member 2[62], which has a pelvis of similar size and morphology as that of the StW 431 male skeleton. Thus, it seems that StW 573 and StW 431 represent, respectively, female and male *A. prometheus*, while the much smaller female pelvis, Sts 14, seems to be that of *A. africanus*. In addition, the femur of *A. prometheus*, as exemplified by StW 573, possesses a short neck joined to a platymeric diaphysis[63], leading to the logical conclusion that the Jacovec femur, StW 598, with a long neck and rounded diaphysis (as described above), probably represents *A. africanus*. Last, both these femoral forms differ from presumptive *Paranthropus* femora from Swartkrans and that of StW 99, from Sterkfontein Member 4.[14,30] Collectively, these results seem to indicate significant interspecific hominin postcranial morphological variability in late Pliocene and early Pleistocene South Africa.

Acknowledgements

Our work was facilitated by a Kellet Mid-Career Award from the University of Wisconsin-Madison to T.R.P. and by a National Research Foundation (South Africa) grant to D.S., and by many years of support from the Palaeontological Scientific Trust (South Africa). In addition, we thank Kathleen Kuman for her consistent financial support of our research through National Research Foundation grants. Kuman also continues to provide logistical and personal support of our research, as do our colleagues, Andrew Phaswana and Abel Molepolle. Thanks to Matt Caruana for general support of our work and especially for some images used in this paper. We thank Bernhard Zipfel for his interest in our research and for facilitating our comparative work at the University of the Witwatersrand. Likewise, we owe Marion Bamford and the Evolutionary Studies Institute at Wits much gratitude for supporting our research at Sterkfontein. Last, we are especially grateful to Jane Carruthers and Margaret Avery for their forbearance and level-headed advice as they shepherded this paper to publication.

Competing interests

We declare that there are no competing interests.

Authors' contributions

T.R.P. conceived and designed the study, collected and analysed the data, and wrote the manuscript. J.L.H. conceived and designed the study and collected and analysed the data. R.J.C. conceived and designed the study and collected and analysed the data. D.S. collected and analysed the data. A.J.H. collected and analysed the data.

References

1. Broom R, Schepers GWH. The South African fossil ape-men: The Australopithecinae. Pretoria: Transvaal Museum; 1946.

2. Tobias PV, Copley K, Brain CK. South Africa. In: Oakley K, Campbell B, Molleson TI, editors. Catalogue of fossil hominids, part I: Africa (revised edition). London: British Museum of Natural History; 1977. p. 95–151.

3. Brain CK. The hunters or the hunted? An introduction to African cave taphonomy. Chicago, IL: University of Chicago Press; 1981.

4. Pickering TR, Clarke RJ, Moggi-Cecchi J. Role of carnivores in the accumulation of the Sterkfontein Member 4 hominid assemblage: A taphonomic reassessment of the complete hominid fossil sample (1936–1999). Am J Phys Anthropol. 2004;125(1):1–15. https://doi.org/10.1002/ajpa.10278

5. Clarke RJ. *Australopithecus* from Sterkfontein Caves, South Africa. In: Reed K, Fleagle J, Leakey REF, editors. Paleobiology of *Australopithecus.* Dordrecht: Springer; 2013. p. 105–123.

6. Heaton JL, Pickering TR, Clarke RJ. Description and initial analysis of hominin craniodental remains from Sterkfontein Members 4-5 and Jacovec Cavern, South Africa (1998–2017). J Hum Evol. In press 2021.

7. Pickering TR, Heaton JL, Clarke RJ, Stratford D. Hominin vertebrae and upper limb bone fossils from Sterkfontein Caves, South Africa (1998–2003 excavations). Am J Phys Anthropol. 2018;3(168):459–480. https://doi.org/10.1002/ajpa.23758

8. Pickering TR, Heaton JL, Clarke RJ, Stratford D. Hominin hand bone fossils from Sterkfontein Caves, South Africa (1998–2003 excavations). J Hum Evol. 2018;1(118):89–102. https://doi.org/10.1016/j.jhevol.2018.02.014

9. Vrba ES. The significance of bovid remains as indicators of environment and predation patterns. In: Behrensmeyer AK, Hill AP, editors. Fossils in the making. Chicago, IL: University of Chicago Press; 1980. p. 247–271.

10. Vrba ES. Biostratigraphy and chronology, based on Bovidae, of southern hominid-associated assemblages: Makapansgat, Sterkfontein, Taung, Kromdraai, Swartkrans; also Elandsfontein (Saldanha), Broken Hill (now Kabwe) and Cave of Hearths. Congrès international de paléontologie humaine, 1er congrès, tome 2. Nice: CRNS; 1982.

11. Delson E. Chronology of South African australopith site units. In: Grine F, editor. Evolutionary history of the 'robust' australopithecines. New York: Aldine de Gruyter; 1988. p. 317–324.

12. McKee JK, Thackeray JF, Berger LR. Faunal seriation of southern African Pliocene and Pleistocene fossil deposits. Am J Phys Anthropol. 1995;3(96):235–250. https://doi.org/10.1002/ajpa.1330960303

13. Kuman K, Clarke RJ. Stratigraphy, artefact industries and hominid associations for Sterkfontein, Member 5. J Hum Evol. 2000;6(38):827–847. https://doi.org/10.1006/jhev.1999.0392

14. Partridge T, Granger DE, Caffee MW, Clarke RJ. Lower Pliocene hominid remains from Sterkfontein. Science. 2003;300:607–612.

15. Mavuso S. The sedimentology of the Jacovec Cavern, Sterkfontein [MSc thesis]. Johannesburg: University of the Witwatersrand; 2018.

16. Kibii J. Comparative taxonomic, taphonomic and palaeoenvironmental analysis of 4–2.3 million-year-old australopithecine cave infills at Sterkfontein [PhD thesis]. Johannesburg: University of the Witwatersrand; 2006.

17. Clarke R. *Australopithecus* and early *Homo* in southern Africa. In: Delson E, editor. Ancestors: The hard evidence. New York: Alan R. Liss; 1985. p. 171–177.

18. Clarke R. Advances in understanding the craniofacial anatomy of South African early hominids. In: Corruccini R, Ciochon R, editors. Integrative paths to the past: Essays in honor of F. Clark Howell. Upper Saddle River, NJ: Prentice Hall; 1994. p. 205–222.

19. Clarke R, Kuman K. The skull of StW 573, a 3.67 Ma *Australopithecus prometheus* skeleton from Sterkfontein Caves, South Africa. J Hum Evol. 2019;134:102634. https://doi.org/10.1016/j.jhevol.2019.06.005

20. Behrensmeyer AK. Taphonomic and ecologic information from bone weathering. Paleobiology. 1978;4(2):150–162. https://doi.org/10.1017/S0094837300005820

21. Blumenschine RJ, Marean CW, Capaldo S. Blind tests of inter-analyst correspondence and accuracy in the identification of cut marks, percussion marks, and carnivore tooth marks on bone surfaces. J Archaeol Sci. 1996;23(4):493–507. https://doi.org/10.1006/jasc.1996.0047

22. Villa P, Mahieu E. Breakage patterns of human long bones. J Hum Evol. 1991;21(1):27–48. https://doi.org/10.1016/0047-2484(91)90034-S

23. Pickering TR, Domínguez-Rodrigo M, Egeland CP, Brain CK. The contribution of limb bone fracture patterns to reconstructing early hominid behaviour at Swartkrans Cave (South Africa): Archaeological application of a new analytical method. Int J Osteoarchaeol. 2005;15(4):247–260. https://doi.org/10.1002/oa.780

24. Alcántara García V, Barba Egido R, Barral del Pino JM, Crespo Ruiz A, Eiriz Vidal A, Falquina Aparicio Á, et al. Determinación de procesos de fractura sobre huesos frescos: un sistema de análisis de los ángulos de los planos de fracturación como discriminador de agentes bióticos [Determination of fracture processes on fresh bones: A system for analyzing the angles of the fracturing planes as a discriminator of biotic agents]. Trabajos de Prehistoria. 2006;63:37–45. Spanish.

25. Martin R. Lehrbuch der Anthropologie in Systematischer Darstellung mit Besonderer Berücksichtigung der Anthropologischen Methoden für Studierende, Ärzte und Forschungsreisend [Textbook of anthropology in a systematic representation with special consideration of anthropological methods for students, doctors and explorers]. Vol. 2. Stuttgart: Gustav Fischer Verlag; 1928. German.

26. Martin R, Saller K. Lehrbuch der Anthropologie in Systematischer Darstellung mit Besonderer Berücksichtigung der Anthropologischen Methoden für Studierende, Ärzte und Forschungsreisend [Textbook of anthropology in a systematic representation with special consideration of anthropological methods for students, doctors and explorers]. Vol. 1. Stuttgart: Gustav Fischer Verlag; no date. German.

27. McHenry HM, Corruccini RS. The femur in early human evolution. Am J Phys Anthropol. 1978;49(4):473–488. https://doi.org/10.1002/ajpa.1330490407

28. Marchi D, Walker CS, Wei P, Holliday TW, Churchill S, Berger LR, et al. The thigh and leg of *Homo naledi*. J Hum Evol. 2017;104:174–204. https://doi.org/10.1016/j.jhevol.2016.09.005

29. Holtby J. Some indices and measurements of the modern femur. J Anat. 1918;52:363–382.

30. Robinson JT. Early hominid posture and locomotion. Chicago, IL: University of Chicago; 1972.

31. Tardieu C. Morpho-functional analysis of the articular surfaces of the knee joint in primates. In: Chiarelli SB, Corruccini RS, editors. Primate evolutionary biology. Berlin: Springer-Verlag; 1981. p. 68–80.

32. Pickering TR, Heaton JL, Zwodeski S, Kuman K. Taphonomy of bones from baboons killed and eaten by wild leopards in Mapungubwe National Park, South Africa*.* J Taphonomy. 2011;9:117–159.

33. Berger LR, Tobias PV. A chimpanzee-like tibia from Sterkfontein, South Africa and its implications for the interpretation of bipedalism in *Australopithecus africanus*. J Hum Evol. 1996;30(4):343–348. https://doi.org/10.1006/jhev.1996.0027

34. Organ J, Ward CV. Contours of the hominoid lateral tibial condyle with implications for *Australopithecus.* J Hum Evol. 2006;51(2):113–127. https://doi.org/10.1016/j.jhevol.2006.01.007

35. Tardieu C. The knee joint in three hominoid primates: Application to Plio-Pleistocene hominids and evolutionary implications. In: Taub D, King F, editors. Current perspectives in primate biology. New York: Van Nostrand; 1986. p. 182–190.

36. Dugan J, Holliday TW. Utility of the lateral meniscal notch in distinguishing hominin taxa. J Hum Evol. 2009;57(6):773–776. https://doi.org/10.1016/j.jhevol.2009.07.006

37. Zipfel B, Berger LR. Partial hominin tibia (StW 396) from Sterkfontein, South Africa. Palaeontol Afr. 2009;44:71–75.

38. Lovejoy CO, Meindl R, Ohman J, Heiple K, White TD. The Maka femur and its bearing on the antiquity of human walking: applying contemporary concepts of morphogenesis to the human fossil record. Am J Phys Anthropol. 2002;119(2):97–133. https://doi.org/10.1002/ajpa.10111

39. Ruff C, Higgins R. Femoral neck structure and function in early hominins. Am J Phys Anthropol. 2013;150(4):512–525. https://doi.org/10.1002/ajpa.22214

40. Lovejoy CO. The natural history of human gait and posture: Part 3, the knee. Gait & Posture. 2007;25(3):325–341. https://doi.org/10.1016/j.gaitpost.2006.05.001

41. DeSilva J, Holt K, Churchill S, Carlson K, Walker C, Zipfel B, et al. The lower limb and mechanics of walking in *Australopithecus sediba*. Science. 2013;340:1232999-1-5. https//doi.org/10.1126/science.1232999

42. Heiple K, Lovejoy CO. The distal femoral anatomy of *Australopithecus*. Am J Phys Anthropol. 1971;35:75–84. https://doi.org/10.1038/235175a0

43. Lovejoy CO, Johanson DC, Coppens Y. Hominid lower limb bones recovered from the Hadar formation: 1974–1977 collections. Am J Phys Anthropol. 1982;57(4):679–700. https://doi.org/10.1002/ajpa.1330570411

44. Jungers WL. Lucy's limbs: Skeletal allometry and locomotion in *Australopithecus afarensis*. Nature. 1982;297:676–678. https://doi.org/10.1038/297676a0

45. Stern JT, Susman RL. The locomotor anatomy of *Australopithecus afarensis*. Am J Phys Anthropol. 1983;60:279–317.

46. Harmon EH. The shape of the early hominin proximal femur. Am J Phys Anthropol. 2009;139(2):154–171. https://doi.org/10.1002/ajpa.20966

47. Lovejoy CO. The natural history of human gait and posture: part 1, spine and pelvis. Gait & Posture. 2005;21(1):95–112. https://doi.org/10.1016/s0966-6362(04)00014-1

48. White TD, WoldeGabriel G, Asfaw B, Ambrose S, Beyene Y, Bernor RL, et al. Asa Issie, Aramis and the origin of *Australopithecus*. Nature. 2006;440:883–889. https://doi.org/10.1038/nature04629

49. Lovejoy CO, McCollum M. Spinopelvic pathways to bipedality: why no hominids ever relied on a bent-hip-bent-knee gait. Philos Trans R Soc Lond. 2010;365:3289–3299. https://doi.org/10.1098/rstb.2010.0112

50. Gracovetsky SA, Iacono S. Energy transfer in the spinal cord. J Biomed Engineer. 1987;9:99–114.

51. Vialle R, Levassor N, Rillardon L, Templier A, Skalli W, Guigui P. Radiographic analysis of the sagittal alignment and balance of the spine in asymptomatic subjects. J Bone Joint Surg A. 2005;87:260–267.

52. Susman RL. Comparative and functional morphology of hominoid fingers. Am J Phys Anthropol. 1979;50(2):215–236. https://doi.org/10.1002/ajpa.1330500211

53. Rein TR. The correspondence between proximal phalanx morphology and locomotion: Implications for inferring the locomotor behavior of fossil catarrhines. Am J Phys Anthropol. 2011;46(3):435–445. https://doi.org/10.1002/ajpa.21599

54. Vrba ES. A new study of the scapula of *Australopithecus africanus* from Sterkfontein. Am J Phys Anthropol. 1979;51:117–130.

55. Clarke RJ, Tobias PV. Sterkfontein Member 2 foot bones of the oldest South African hominid. Science. 1995;269:521–524. https://doi.org/10.1126/science.7624772

56. Macchiarelli R, Bondioli L, Galichon V, Tobias PV. Hip bone trabecular architecture shows uniquely distinctive locomotor behaviour in South African australopithecines. J Hum Evol. 1999;36(2):211–232. https://doi.org/10.1006/jhev.1998.0267

57. Kibii JM, Clarke RJ. A reconstruction of the StW 431 *Australopithecus* pelvis based on newly discovered fragments. S Afr J Sci. 2003;99:225–226.

58. Haeusler M. New insights into the locomotion of *Australopithecus africanus* based on the pelvis. Evol Anthropol Suppl. 2002;1(S1):53–57. https://doi.org/10.1002/evan.10056

59. McHenry HM, Berger LR. Body proportions in *Australopithecus afarensis* and *A. africanus* and the origins of the genus *Homo*. J Hum Evol. 1998;35:1–22.

60. Harmon E. Size and shape variation in the proximal femur of *Australopithecus africanus.* J Hum Evol. 2009;56(6):551–559. https://doi.org/10.1016/j.jhevol.2009.01.002

61. Deloison Y. Anatomie des os fossils de pieds des hominids d'Afrique du sud dates entre 2.4 et 3.5 millions d'annés: interprétation quant à leur mode de locomotion [Anatomy of the fossil bones of the feet of South African hominids between 2.4 and 3.5 million years ago: Interpretation of their mode of locomotion]. Biométrie Hum Anthropol. 2003;21:189–230. French.

62. Clarke R. Excavation, reconstruction and taphonomy of the StW 573 *Australopithecus prometheus* skeleton from Sterkfontein Caves, South Africa. J Hum Evol. 2019;127:41–53. https://doi.org/10.1016/j.jhevol.2018.11.010

63. Heaton J, Pickering TR, Carlson K, Crompton R, Jashashvili T, Beaudet A, et al. The long limb bones of the StW 573 *Australopithecus* skeleton from Sterkfontein Member 2: Descriptions and proportions. J Hum Evol. 2019;133:167–197. https://doi.org/10.1016/j.jhevol.2019.05.015

64. Reynolds SC, Kibii JM. Sterkfontein at 75: Review of palaeoenvironments, fauna and archaeology from the hominin site of Sterkfontein (Gauteng Province, South Africa). Palaeontol Afr. 2011;46:59–88.

65. Clarke RJ. A deeper understanding of the stratigraphy of the Sterkfontein fossil hominid site. Trans R Soc S Afr. 2006;61(2):11–120. https://doi.org/10.1080/00359190609519960

66. Robinson JT. Australopithecines and artefacts at Sterkfontein, part I: Sterkfontein stratigraphy and the significance of the Extension Site. S Afr Archaeol Bull. 1962;17(66):87–107. https://doi.org/10.2307/3886942

67. Pickering TR, Heaton J, Clarke R, Sutton M, Brain CK, Kuman K. New hominid fossils from Member 1 of the Swartkrans Formation, South Africa. J Hum Evol. 2012;62(5):618–628. https://doi.org/10.1016/j.jhevol.2012.02.003

68. Tardieu C, Glard Y, Garron E, Boulay C, Jouve J-L, Dutour O, et al. Relationship between formation of the femoral bicondylar angle and trochlear shape: Independence of diaphyseal and epiphyseal growth. Am J Phys Anthropol. 2006;130(4):491–500. https://doi.org/10.1002/ajpa.20373

The cryptic case of *Otomys sloggetti* (Sloggett's vlei rat): Interpreting murid molar morphology in the fossil record

AUTHORS:
Thalassa Matthews[1]
Turid H. Nel[2]

AFFILIATIONS:
[1]Centre of Excellence in Palaeosciences, Iziko Museums of South Africa, Cape Town, South Africa
[2]SFF Centre for Early Sapiens Behaviour (SapienCE), University of Bergen, Bergen, Norway

CORRESPONDENCE TO:
Thalassa Matthews

EMAIL:
tmatthews.matthews@gmail.com

DATES:
Received: 10 Sep. 2019
Revised: 01 June 2020
Accepted: 17 July 2020
Published: 29 Jan. 2021

HOW TO CITE:
Matthews T, Nel TH. The cryptic case of *Otomys sloggetti* (Sloggett's Vlei Rat): Interpreting murid molar morphology in the fossil record. S Afr J Sci. 2021;117(1/2), Art. #7137. https://doi.org/10.17159/sajs.2021/7137

ARTICLE INCLUDES:
☒ Peer review
☐ Supplementary material

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

EDITOR:
Margaret Avery

KEYWORDS:
Otomyinae, Pinnacle Point, Klipdrift Shelter, tooth morphology

FUNDING:
South African National Research Foundation, National Research Council of Norway

Vlei rats (Family: Muridae; Subfamily: Otomyinae) have a widespread distribution in southern Africa. They are favoured prey of barn and spotted eagle owls, and frequently become associated with archaeological deposits when the owls roost in cave sites. The phylogeny of several Otomyinae species is enigmatic, and *Otomys sloggetti* (Sloggett's vlei rat) is no exception. This species has been referred to as the 'ice rat' and present distribution ranges are seemingly limited to mountainous areas, at high altitude, in Lesotho, Drakensberg and the Karoo. It was thus surprising and unexpected when specimens closely resembling *Otomys sloggetti* (identification was based on molar morphology) were found in several archaeological sites on the south and west coasts of South Africa, and also in modern owl pellet assemblages – all extralimital to the current reported distribution. However, further examination of and comparison between these specimens, as well as extensive differences observed between comparative *Otomys sloggetti* specimens from museum collections, highlighted potential problems associated with the common practice of using tooth morphology to identify fossil murid species. We identified six molar morphotypes from the fossil and modern material, all of which bore a morphological resemblance to *O. sloggetti*. The material discussed in this paper suggests that cryptic, undescribed vlei rat species, or subspecies, have been in the past, and may yet be, co-occurring with modern populations of *O. karoensis* and *O. irroratus*. Phylogenetic studies need to be done in conjunction with morphological studies, as, currently, the relationship between the huge variation seen in interspecific morphology with genetics is little understood, different *Otomys* species are not always distinguishable morphologically, and considerable chromosomal polytypes have been found. Our findings highlight the need for extensive cladistic and genetic research on the Otomyinae.

Significance:

- Mice and shrews from fossil sites are frequently used by archaeologists as indicators of past climatic and environmental conditions. Research into the species present in fossil assemblages is usually done on a single site basis and intersite comparisons are rare. The taxonomic conundrums presented by a vlei rat found in several South African archaeological sites indicates that such comparisons could result in the re-evaluation of identifications, and/or indicate the presence of cryptic species/subspecies. Phylogenetic studies are needed in conjunction with morphological studies, as the relationship between variations in interspecific tooth morphology (used to identify taxa) with genetics is little understood. This in turn will help to elucidate the relationship between morphology, biogeography and local adaptations.

Introduction

The identity, and biogeographical affiliations, of fossil rats and mice (Family: Muridae) are important as these small animals, which have short lifespans, small ranges, and in some cases particular habitat requirements, are used to provide environmental information in order to reconstruct ancient environments, and trace palaeoenvironmental change over time. The correct identification of these animals in fossil sites is thus important as errors can lead to erroneous assumptions and conclusions being made about past environments, and the effects of climate change. Here we focus on a vlei rat species, namely *Otomys sloggetti,* and use the fossil record of this taxon on the south coast of South Africa to illustrate some of the issues and potential problems associated with the common practice of using tooth morphology to identify fossil murid species.

Research into the murid species present in fossil assemblages is usually done on a single site basis, and it is extremely rare for researchers to compare the same taxon across sites. The taxonomic conundrums presented in this paper indicate that such comparisons might result in the re-evaluation of identifications, as inconsistencies and errors may ensue if museum specimens have been incorrectly identified – a scenario which is particularly likely in the case of cryptic species (species which look morphologically the same, but belong to different species). Cross-site comparisons, such as those made here, may also indicate variability within a species suggestive of misidentification, and/or the presence of cryptic species. We compare the molar morphology of a variety of fossil *Otomys* cf. *sloggetti* from a number of south coast archaeological sites, and a west coast site, with each other, with modern *Otomys sloggetti*, and with two other *Otomys* species (*O. irroratus* and *O. karoensis*). These comparisons include comparative material from museum collections, and from two south coast barn owl pellet collections.

Taxonomy and phylogeny of Otomys sloggetti

Otomys sloggetti is a rat which belongs to the relatively speciose 'vlei rat' (also called 'laminate toothed rat') tribe. These herbivorous short-tailed rodents (Subfamily: Murinae) are frequently mentioned as being 'vole-like' in appearance. The Otomyini are represented by some 32 recognised species (including two of which are undescribed) and are endemic and widespread in sub-Saharan Africa, with a number of species occurring in South Africa.[1] The exact number of extant *Otomys* species is contentious and a number of cryptic species have been identified, while others await further elucidation.[2-5] Some studies suggest that speciation has occurred along biomes, with phenotypic and genotypic divergent lineages corresponding to the Fynbos/Albany Thicket and Grassland biomes in the case of *Otomys irroratus* and *Otomys karoensis*.[2,3] Geometric morphometric analyses carried out on the crania of *Otomys unisulcatus* also failed to support the genetic groupings, but rather followed biome boundaries, indicating previous environmental adaptations.[6] This, together with the adaptability observed in fossil *Otomys* communities when faced with environmental and climatic change on the south coast of South Africa[7], as well as the fact that up to three or more *Otomys* species may live sympatrically in an area[1,2,7], indicates the flexibility and plasticity of this murid subfamily.

The phylogeny of *Otomys sloggetti*, and the number of lineages and species present, as is the case with many of the other otomyine species, is currently unresolved.[1,2] In a study that looked at mitochondrial and nuclear genes, as well as morphological characteristics, specimens previously identified as *Otomys sloggetti* were found to represent two clades, and three lineages.[8] A cladistic analysis of 45 morphological (craniodental) and 46 binary allozyme characters indicated that *Otomys sloggetti* was basal to a 'mesic clade' of southern and eastern African species[9], and the idea that *Otomys sloggetti* represents a basal lineage has been reiterated elsewhere[10]. The most recent research into the phylogeny of *O. sloggetti* by Taylor et al.[2] differentiated three distinct mtDNA clades of *O. sloggetti* in the Drakensberg, Maluti and Sneeuberg ranges, respectively, extending the range of this species to beyond the Sneeuberg and the southern Great Escarpment. Further morphometric and karotypic analyses are required to elucidate the extent to which these species are cryptic.[2] The five laminae observed in the upper third molar of *Otomys sloggetti,* and also in the extinct fossil species *Otomys gracilis*, are noted as being a feature of earlier forms of the genus, and it is suggested that younger (more recent) taxa, such as *O. irroratus, O. karoensis* and *O. laminatus,* have more laminae.[11] Previously, *O. sloggetti* and *O. unisulcatus* were placed in a separate family (*Myotomys*)[12], but recent genetic research has resulted in them being reclassified as *Otomys*[1,2]. Fossil *Otomys sloggetti* have been found in a few Cradle of Humankind fossil sites in Gauteng Province, including Gladysvale Cave[13], Border Cave[14] (where cf. *O. sloggetti* was tentatively identified) and Sterkfontein Cave[15], which date as far back as 3.3 Ma[16].

Small mammals as palaeoenvironmental indicators

As mentioned previously, archaeologists and palaeontologists frequently use rats, mice and shrews from fossil sites as indicators of past climatic and environmental conditions.[17-19] These animals generally become associated with fossil deposits when barn and eagle owls, which roost in the same caves occupied by people, regurgitate the bones and teeth of their prey in their pellets. Over time, the pellets disintegrate, leaving lenses of bones and teeth in the sediment. As these owls have small home ranges, and select a broad range of prey, they provide a good sample of the small mammal species living in the vicinity of the cave site.[14,17] The teeth of rodents and soricids (shrews) are generally used to identify the small mammal species present in archaeological deposits. The laminate molars of vlei rats, and the terminology used in this paper, is illustrated in Figure 1.

Vlei rats have laminate teeth and generally show a high degree of intraspecific variability in terms of molar shape, and sometimes even in the number of laminae. For example, *O. karoensis* is noted as sometimes having five, rather than six, laminae on the upper third molar.[1] This variability, which includes biogeographical differences in morphology and size, and the fact that the appearance of the occlusal surface of the tooth changes over time with wear (in some species wear leads to the joining of previously separate laminae), makes this a difficult group of rodents to identify. In fossil deposits, the cranial bones are frequently broken to the degree that only isolated molars are recovered. This means that the approach of identifying features used with modern comparative material, such as presence/absence of grooves on the lower incisor, and the shape of the petrotympanic foramen, cannot be used. *O. sloggetti* is distinguishable from all other *Otomys* species in having ungrooved lower incisors, and a slit-shaped petrotympanic foramen.[2,9] In the case of material from fossil deposits, however, the lower first molar (M_1) of the mandible, and upper third molar (M^3) of the maxilla (see Figure 1), are the most distinctive and useful teeth for the identification of *Otomys* species.

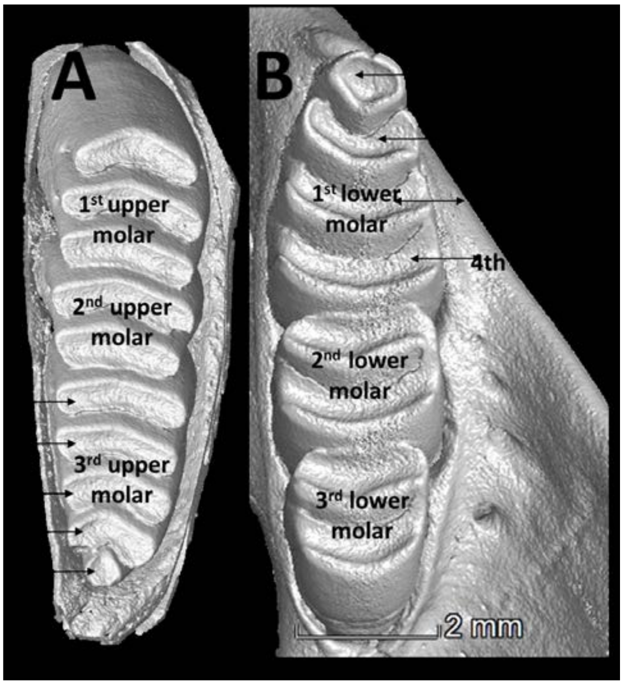
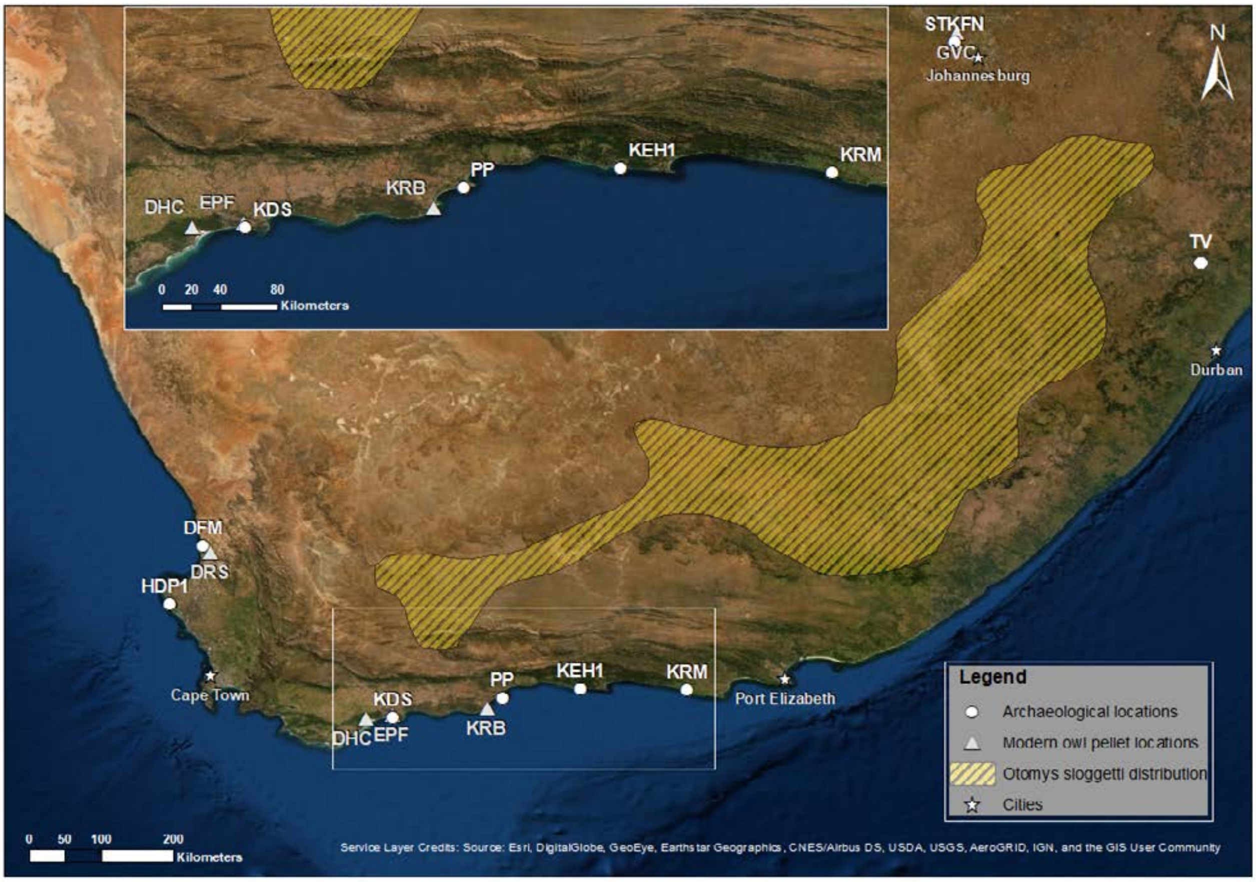


Figure 1: Terminology used for the upper and lower laminate molar teeth of vlei rats (*Otomys* sp.): (A) right maxillary tooth row of *O. sloggetti* and (B) right mandibular tooth row of *O. sloggetti* (Ditsong National Museum specimen TM–22669).

The fossil specimens referred to in this article are currently curated at the Iziko South African Museum and come from a number of archaeological sites including the Pinnacle Point complex (PP), Knysna Heads 1 (KEH1), Dunefield Midden (DFM), Sterkfontein Cave (STKFN), Klasies River main site (KR) and Klipdrift Shelter (KDS) (see Figure 2 for these localities). Comparative, modern material was obtained from the Ditsong National Museum of Natural History, as well as Iziko. Modern *Otomys* specimens were recovered from owl pellet assemblages collected at Elandspad Farmhouse (EPF) and De Hoop Collections Cool Room (DHC) in the De Hoop Nature Reserve, Keurbooms (KRB) in the Wilderness area, and Diepkloof Rock Shelter (DRS) on the west coast. Figure 2 shows the fossil and pellet collection sites mentioned in this paper, and the current extent of *O. sloggetti*'s distribution based on the IUCN Red List.[20] Computed tomography (CT) scans of individual fossils, and some of the modern material, were carried out at the Central Analytical Facilities of Stellenbosch University (South Africa), using their micro- and nano-CT scanning facilities. Length and breadth measurements (see Appendix 1) of all M_1 and M^3 molars were made on nano-CT scans of the specimens, using the 'measuring' and 'dimensions' functions of VGMax, version 3.3. Measurements of all the teeth presented here are given in Appendix 1. Note that these must be considered within the context of age and wear patterns, as older individuals, with more worn teeth, will automatically yield higher measurements than younger individuals, as the occlusal surface becomes longer, and broader, with wear. Photographs of modern specimens from EPF were taken with a Leica M205A camera attached to a Leica M275 microscope. Specimens from KDS were photographed with a Leica DFC 295 camera attached to a Leica M125 C microscope.



DFM, Dunefield Midden; DHC, De Hoop Collections Cool Room; DRS, Diepkloof Rock Shelter; EPF, Elandspad Farmhouse; GVC, Gladysvale Cave; HDP1, Hodjiespunt 1; KDS, Klipdrift Shelter; KEH1, Knysna Heads 1; KRM, Klasies River; PP, Pinnacle Point; STKFN, Sterkfontein Cave; TV, Tugela Valley

Figure 2: Localities of owl pellet collections and fossil sites.

Otomys sloggetti *and the south coast fossil record*

Otomys sloggetti sensu lato frequently co-occurs with two other *Otomys* species, namely *Otomys karoensis* and *Otomys irroratus,* in fossil deposits on the south coast.[7] *O. sloggetti* sensu lato was identified in the fossil deposits based on the M^3 and M_1, and is differentiated by two specific molar morphologies. Firstly, by the presence of five laminae on the upper third molar (M^3) as opposed to the six laminae generally found in *Otomys karoensis* and *Otomys irroratus*, and secondly, a rounded anterior first laminae (M_1) (which differs to the rectangular first lamina of the other two species), and in the distribution of the laminae across the occlusal surface of the molar. Another vlei rat with the same general morphology of the M_1 and M^3, *O. unisulcatus*, has also made an extralimital appearance in some south coast fossil horizons, but the shape and general appearance of the molars are differentiable from that of *O. sloggetti* (see Figures 5 and 7) as the M^3 has generally smaller dimensions and is roughly rectangular in shape, and the anterior and second laminae of the M_1 merge to form a characteristic 'C' shape, with only slight wear.

The appearance of *O. sloggetti* in a number of fossil south coast South African sites (see Figure 2 for localities) was unexpected, as it has been called a high altitude and alpine endemic species, and its current distribution includes the Drakensberg, Sneeuberg and the southern Great Escarpment[2], with a couple of isolated populations on mountains in the Karoo[8,21-23]. This species is often referred to as the 'ice rat' due to its appearance in areas which reach icy temperatures. The appearance of *O. sloggetti* in the south coast fossil record was surprising as it appeared to belie the categorisation of this taxon as being a 'high altitude endemic' and 'montane-adapted'[9]. *Otomys sloggetti* was identified in three coastal cave sites situated on Pinnacle Point, just to the southwest of Mossel Bay (PP9C, PP30 and PP13B), as well as in another coastal archaeological cave site situated north of the Pinnacle Point sites on the eastern Knysna Head (KEH1).[7] At KEH1, *O. sloggetti* was identified in horizons dating to around the Last Glacial Maximum (~23 ka), and in the Pinnacle Point sites sometime during the period 157–180 ka (PP13B), at ~150 ka (PP30), and at ~130 ka (PP9C).[7] Due to an observed high degree of variation between fossil specimens, existing phylogenetic uncertainties, and variation between modern *O. sloggetti* teeth obtained from comparative museum collections, identification of the fossil taxa was later changed to *O.* cf. *sloggetti* sensu lato. The appearance of *O.* cf. *sloggetti* in surface deposits in PP9C and PP13C was suggested to have resulted from the mixing of surface and fossil deposits, and/or be an indication of the persistence of this taxon until fairly recently in the region.[7]

Further southwards, additional specimens (*O. sloggetti* sensu lato) were recovered from Klipdrift Shelter, a coastal archaeological site situated in the De Hoop Nature Reserve. Two specimens were found in horizons dated to 51.7±3.3 ka, and another in deposits dated to 60.3±3.8 ka.[24] At least three individuals were found in layers that are undated, but assumed, on the basis of stratigraphy, to be younger than 52 ka. *Otomys sloggetti* sensu lato (represented by M_1 molars) was also recovered from the Klasies River main site in the Howiesons Poort layers (~65–46 ka) (T.H.N. unpublished data), and has also been identified in a couple of west coast fossil sites dating to the Holocene (Dunefield Midden) and Late Pleistocene (at the palaeontological site of Hoedjiespunt 1[25]).

The unexpected find of what appeared to be *O.* cf. *sloggetti* in fossil deposits was then repeated during investigations of micromammals from modern owl pellet accumulations (Figures 3 and 4). Two five-laminate M^3 molars were recovered from a comparative eagle owl pellet collection made from Keurbooms, near Wilderness on the south coast in 2005[7], and additional *O.* cf. *sloggetti* teeth were found by T.H.N., in October 2018, in barn owl pellets from two separate locations in the De Hoop Nature Reserve. On the west coast, *O.* cf. *sloggetti* was also found (represented by M_1 and M^3 molars) in what appeared to be a relatively modern, disaggregated barn owl roost site from Diepkloof Rock Shelter near Elands Bay, which had been deposited on top of an archaeological horizon (Parkington JP 2018 June 1, personal communication).

As mentioned above, *O.* cf. *sloggetti* was identified on the basis of five laminae on the M^3, and four laminae on the M^1 (with the anterior laminae of the latter being roughly ovoid in shape, and clearly differing in morphology from associated *O. karoensis* and *O. irroratus* specimens which have rectangular-shaped anterior laminae). This diagnosis became more complicated, however, as morphological differences in the cranial bones and the M^3 were then noted between two incomplete skulls obtained from modern owl pellet collections made in De Hoop Nature Reserve, namely the Elandspad Bulk Pellet sample (EPF), and De Hoop Collections Cool Room roost site (DHC). This find had potential implications for the identification of *O. sloggetti* in the fossil record on the basis of using a five-laminate M^3, and it became clear that further investigation was needed into the identity of the modern five-laminate taxa. The two skulls were thus CT scanned and compared with those of *O. irroratus*, *O. karoensis*, and *O. sloggetti* (Figures 3 and 4). The former two species were included as they are ubiquitous throughout South African fossil sites.[6]

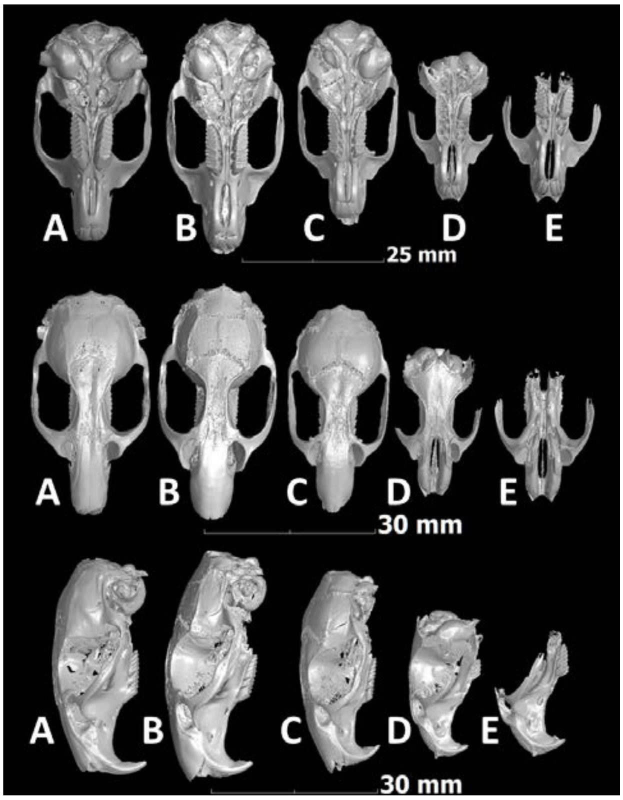


Figure 3: Comparison between the skulls of *Otomys sloggetti*, *O. irroratus, O. karoensis,* and two *Otomys* taxa from modern owl pellet collections in ventral (top), dorsal (middle) and left lateral (bottom) view: (A) *O. sloggetti* (TM–16519B), (B) *O. irroratus* (ZM–82140), (C) *O. karoensis* (ZM–37374), (D) *Otomys* sp. De Hoop Collections Cool Room roost site (DHC–10–3), (E) *Otomys* sp. Elandspad bulk pellet sample (EPF–bulk–1).

Figure 3 illustrates that *O. sloggetti* (TM–16519B) differs quite considerably from the two skulls from the pellets, as well as *O. karoensis* and *O. irroratus*, in general skull form and proportions: the skull cap is rounder in shape relative to the others, the orbital foramen is broader and extends less anteriorly to posteriorly, the nasal foramen is rounded in shape rather than rectangular, and the shape and size of the maxilla, zygomatic arch and the zygomatic plate are likewise clearly different. The upper incisors of all the taxa exhibit a single groove. The DHC–10–3 skull appears very similar, although not identical, to *O. karoensis* in terms of the shape of the zygomatic arch, general skull shape, the infraorbital foramen, and the premaxilla. Small differences may be attributed to the fact that it belongs to a smaller and (as indicated by the wear pattern on the molars) younger, individual. The premaxilla and maxilla from the Elandspad Bulk Sample (EPF–bulk–1) specimen differed in size and morphology, not only to *O. sloggetti*, but also to all the other taxa illustrated in Figure 3. In order to illustrate these morphological differences, and to compare this specimen with another of a similar M^3 molar morphology, Figure 4 focuses on the maxilla and the zygomatic arch of EPF–bulk–1 and compares it with the modern comparative five-laminate M^3 cf. *O. karoensis* (DHC–10–3).

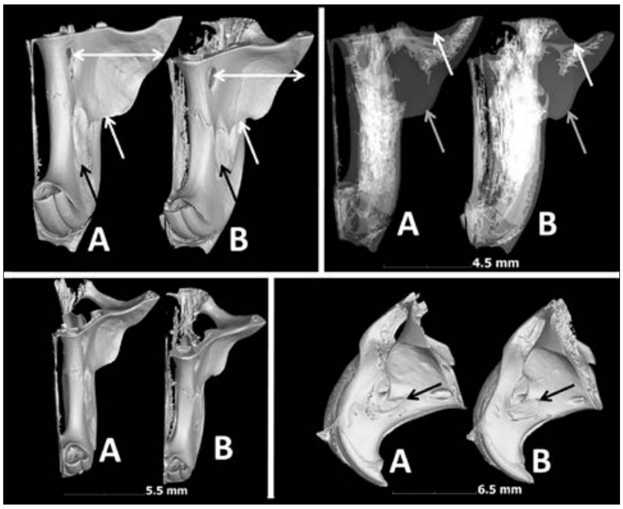
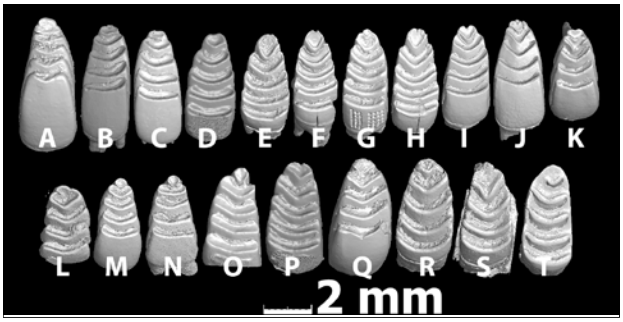


Figure 4: Comparison of maxilla of (A) indeterminate Elandspad *Otomys* (EPF–bulk–1) with (B) cf. *O. karoensis* (DHC–10–3): top left, right maxilla ventral view; top right, view of same with enhanced transparency; bottom left, view of intraorbital foramina; bottom right, left premaxilla. Arrows indicate areas of morphological difference between the two taxa.

The area where the incisive bone meets the zygomatic process of the maxilla is curved in *O. karoensis* (DHC-10-3) but straight in the indeterminate *Otomys* from EPF. Other differences include the size and shape of the infraorbital foramen (much larger in the case of the EPF *Otomys*), and the zygomatic arch is different in size and shape between the taxa. The differences observed in maxillary morphology are likewise reflected in the M^3, as illustrated in Figure 5, where the distribution of the laminae across the occlusal surface of the tooth, and the size and shape of the anterior lamina, differ. The fact that the Elandspad specimen shows morphological differences in the maxilla *and* the M^3 supports the identification of this as a taxon or morphotype differentiable from *O. karoensis*. A search was made through the Elandspad owl pellets for M_1 teeth which looked different from those of *O. karoensis* and *O. irroratus*, and which could possibly be a match for the M^3 Elandspad specimen. Two such specimens were recovered from the Elandspad bulk collection, one of which is illustrated in Figure 7. It is distinguished by a notably large, round, first lamina (EPF-bulk-12, see inset photo in Figure 7).

Comparison between extant and fossil *Otomys* taxa

Figure 5 shows a comparison of the M^3 of modern *O. sloggetti* from two different regions in South Africa with each other, and with fossil specimens from archaeological sites on the south coast, west coast, and the Cradle of Humankind (Sterkfontein Cave, Gauteng Province). An *Otomys irroratus* specimen is included in both Figures 5 and 7 to illustrate the similarity of this species in terms of molar morphology to *O. karoensis*. The site name and specimen reference numbers are provided in brackets in Figure 5.



PP, Pinnacle Point; DHC, De Hoop Collections Cool Room roost site; EPF, Elandspad bulk pellet sample; NC, Name Chamber; TM, Ditsong National Museum; KDS, Klipdrift Rock Shelter; DRS, Diepkloof Rock Shelter; STKFN–NC, Sterkfontein Cave Undifferentiated Name Chamber deposits; DFM, Dunefield Midden

Figure 5: Comparison between modern and fossil *Otomys* specimens (M^3). The site name and reference are provided in brackets.

Top row: (A) Fossil *O. irroratus* (PP30–417892), (B) *O. karoensis* (DHC 10–2), (C) cf. *O. karoensis* (DHC 10–3), (D) indeterminate *Otomys* sp. (EPF–bulk–1), (E) fossil *Otomys* sp. (PP9C–100662), (F) fossil *Otomys* sp. (PP13B–99717), (G) fossil *O. karoensis* (PP9C–1005231), (H) fossil *Otomys* sp. (KDS–CR1145), (I) fossil *Otomys* sp. (KDS–CR1146), (J) fossil *Otomys* sp. (KDS–CR570), (K) fossil *Otomys* sp. (KDS–CR1147).

Bottom row: (L) *O. unisulcatus* (DRS, Bag 8), (M) fossil *O. sloggetti* (STKFN–NC), (N) fossil *O. sloggetti* (STKFN–NC), (O) *O. sloggetti* (TM–16519), (P) *O. sloggetti* (TM–22669), (Q) fossil *Otomys* sp. (PP30–100670), (R) fossil *Otomys* sp. (DFM–JAC 13), (S) fossil *O. Otomys* sp. (DFM–NIC 16), (T) fossil *Otomys* sp. (DFM–NIC 27).

The morphological differences could be attributable to biogeography, but are far greater than the degrees of interspecific variation generally observed in murids (T.M. and T.H.N. personal observation), and most likely reflect genetic divergence, like that observed between modern *O. sloggetti* populations living in mountainous areas in the Drakensberg, Sneeuberg and Lesotho.[2] This suggestion is supported by research[8] which notes that specimens identified as *O. sloggetti* in the field were found to represent two genetic clades. The one clade was consistent with the description of *O. sloggetti,* but the other was distinct, not only from *O. sloggetti*, but from all other *Otomys* species.[2,8] This species, endemic to the southern Drakensberg Mountain Range, has subsequently been described (nov. sp. *Otomys willani*).[2] It is differentiated from *O. sloggetti* by six laminae on the M^3 and grooved lower incisors. Intraspecific variation, albeit to a less marked degree, is also observed in the Sterkfontein Cave *O. sloggetti* specimens (Figure 5, M and N) in terms of the degree of curvature of the laminae, and general tooth shape. Both specimens come from the so-called 'Name Chamber' deposits which are mixed in that they contain deposits from both Member 4 and Member 5E, with the Oldowan (Member 5E) deposits dating to ~2–1 Ma, and/or Member 4 (~2.8–2 Ma).[26] The taxonomic relevance of the morphological differences is unclear, given the stratigraphic/age uncertainties. The Sterkfontein fossil specimens are considerably smaller than the modern comparative *O. sloggetti* and almost all the fossil *O. sloggetti* sensu lato, providing an example of the flexibility in size of this genus in both modern and fossil populations (see Appendix 1).

As mentioned previously, *O. karoensis* has been noted to sometimes have five laminae on the M^3. The DHC specimens in Figure 5 ('B' and 'D') illustrate such a case. These specimens differ slightly in that DHC 10–2 ('B') has a tiny auxiliary cusplet situated proximally on the anterior laminae, which indicates a variation occurring in the phenotype. Such additional cusplets (note this feature also appears in the fossils 'G' and J' in Figure 5) would disappear with wear and would not be discernible in older individuals. This kind of intraspecific variation is not uncommon among murids, and small additional cusplets occur infrequently on murid teeth (T.H.N. and T.M. personal observation).

The EPF M^3 (EPF-bulk-1, see 'D' in Figure 5), whose maxilla and premaxilla show differences to *O. karoensis* as noted above, is different from the fossil and modern *O. karoensis* in that it has a larger and more rounded anterior cusp, which is embedded in the second lamina to a greater degree than seen in *O. karoensis*. It also differs in that the spacing between the anterior to the posterior lamina of the occlusal surface is wider, and the laminae are more curved and orientated centrally towards the middle of the tooth, whereas *O. karoensis* laminae slant towards the labial side of the tooth.

The south coast fossil specimen from PP30 (Figure 5 'Q') shows a close fit in terms of morphology and size with the modern *O. sloggetti* specimen TM–22669 (Figure 5 'P'), as does one of the three fossils from Dunefield Midden (Figure 5 'R'), but the two other Dunefield Midden fossil specimens (Figure 5 'S' and 'T'), also originally identified as *O. sloggetti* sensu lato*,* are harder to match as the occlusal enamel-dentine surface of the teeth is very worn.

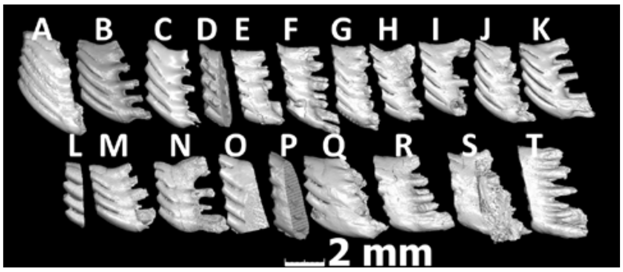
Four fossil specimens from Pinnacle Point (PP9C and PP13B) (Figures 5 and 6, E, F and G), and four from Klipdrift shelter (Figures 5 and 6, H, I, J and K), were previously thought to represent *O.* cf. *sloggetti;* however, they appear morphologically very similar to the modern specimen from EPF (EPF–bulk–1), in terms of size, morphology, orientation of laminae, and general tooth shape. The fossil *Otomys* 'G' (PP13B), as well as 'J' and 'K' (KDS), all have five laminae (G and J also exhibit a tiny auxillary cusplet), and are morphologically very similar to the five-laminate *O. karoensis* in terms of shape and orientation of laminae, and we tentatively suggest that they belong to this species*.*

A comparison of the specimens initially identified as *O. sloggetti* during analysis of the various archaeological sites may be split into two morphological groups, arbitrarily called Morphotype 1 and Morphotype 2. Morphotype 1 (Figure 5: D, E, F, H and I) includes the modern EPF specimen, and fossils from KDS, PP9C and PP13B. Morphotype 1 is defined as exhibiting a large, roundish anterior laminae which is embedded into the second laminae, and the other laminae are curved. This differs to *O. karoensis* and *O. irroratus* where the laminae are more linearly arranged, and the anterior lamina is relatively smaller and more rounded-rectangular in shape. Morphotype 2 was identified in the fossil sites of PP30 and DFM and includes P, Q and R, and probably S and T, as well as one of the modern comparative *O. sloggetti* (TM–22669), which the fossils most closely resemble in terms of general morphology, size, and orientation of laminae. Morphotype 2 teeth yielded greater breadth measurements (breadth of M^3: 2.0–2.2 mm, length of M^3: 3.2–4 mm) than Morphotype 1 teeth (breadth of M^3: 1.8–1.9 mm, length of M^3: 2.9–3.6 mm), but overlapped in terms of length.

When viewed laterally, the five, or six, laminae present on the various *Otomys* taxa, as well as the auxillary cusplets, are clearly discernible (Figure 6). Note that the specimens, and the order of specimen line-up, are identical in Figures 5 and 6.

Figure 7 makes a similar comparison to that of Figure 5, but this time with the M_1 from modern *O. sloggetti*, fossil specimens from the south and west coasts and Sterkfontein Cave. Unfortunately, no M_1 belonging to *O.* cf. *sloggetti* or *O. karoensis* were recovered from Dunefield Midden and this site is thus not represented in Figure 7.

In Figure 7, the top row of specimens from B to E are considered to represent *O. karoensis*, although E from PP9C was originally differentiated from other *O. karoensis* at the site on the basis of size and the relative roundness of the first lamina. However, it appears very similar to the modern specimen from DHC 10–2 (Figure 7, D), and given this likeness, is most probably an *O. karoensis*. Fossil *O. unisulcatus* specimens are represented in Figure 7 (see S, T, U and V), while W represents a modern specimen (from Diepkloof Rock Shelter).

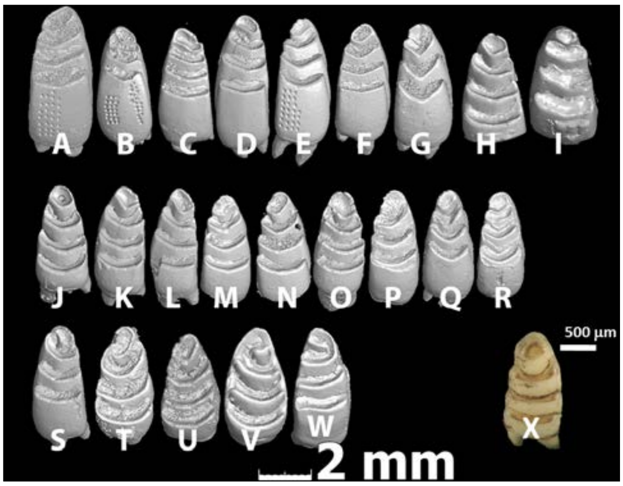


PP, Pinnacle Point; DHC, De Hoop Collections Cool Room roost site; EPF, Elandspad bulk pellet sample; TM, Ditsong National Museum; KDS, Klipdrift Rock Shelter; DRS, Diepkloof Rock Shelter; STKFN–NC, Sterkfontein Cave Undifferentiated Name Chamber deposits; DFM, Dunefield Midden

Figure 6: Comparison in lateral view between modern and fossil *Otomys* specimens (M^3).

Top row: (A) Fossil *O. irroratus* (PP30–417892), (B) *O. karoensis* (DHC 10–2), (C) cf. *O. karoensis* (DHC 10–3), (D) indeterminate *Otomys* sp. (EPF–bulk–1), (E) fossil *Otomys* sp. (PP9C–100662), (F) fossil *Otomys* sp. (PP13B–99717), (G) fossil *O. karoensis* (PP9C–1005231), (H) fossil *Otomys* sp. (KDS–CR1145), (I) fossil *Otomys* sp. (KDS–CR1146), (J) fossil *Otomys* sp. (KDS–CR570), (K) fossil *Otomys* sp. (KDS–CR1147).

Bottom row: (L) *O. unisulcatus* (DRS, Bag 8), (M) fossil *O. sloggetti* (STKFN–NC), (N) fossil *O. sloggetti* (STKFN–NC), (O) *O. sloggetti* (TM–16519), (P) *O. sloggetti* (TM–22669), (Q) fossil *Otomys* sp. (PP30–100670), (R) fossil *Otomys* sp. (DFM–JAC 13), (S) fossil *O. Otomys* sp. (DFM–NIC 16), (T) fossil *Otomys* sp. (DFM–NIC 27).



PP, Pinnacle Point; DHC, De Hoop Collections Cool Room roost site; EPF, Elandspad bulk pellet sample; NC, Name Chamber; TM, Ditsong National Museum; KDS, Klipdrift Rock Shelter; DRS, Diepkloof Rock Shelter; KRM, Klasies River Main; STKFN–NC, Sterkfontein Cave Undifferentiated Name Chamber deposits; DFM, Dunefield Midden; KEH1, Knysna Heads Cave 1

Figure 7: Comparison between fossil and modern *Otomys* species (M_1).

Top row: (A) Fossil *O. irroratus* (PP9C–100431), (B) fossil *O. karoensis* (PP9C–100586), (C) fossil *O. karoensis* (PP30–417855), (D) cf. *O. karoensis* (DHC 10–2), (E) fossil *Otomys* sp. (PP9C–100438), (F) fossil *O. sloggetti* (STKFN–NC), (G) fossil *Otomys* sp. (PP30–418015), (H) *O. sloggetti* (TM–22669), (I) *O. sloggetti* (TM–16519).

Middle row: (J) Fossil *Otomys* sp. (KRM–CR2772), (K) fossil *Otomys* sp. (KRM–CR4776), (L) fossil *Otomys* sp. (KRM–CR4712), (M) fossil *Otomys* sp. (PP13B–99201), (N) fossil *Otomys* sp. (KEH1-PF-4758), (O) fossil *Otomys* sp. (KEH1-PF-9697), (P) fossil *Otomys* sp. (PP30–100654), (Q) indeterminate *Otomys* sp. (DRS, Bag 8), (R) fossil *Otomys* sp. (KDS-CR685).

Bottom row: (S) Fossil *O. unisulcatus* (PP30–418027), (T) fossil *O. unisulcatus* (PP9C–100534B), (U) fossil *Otomys* sp. (PP13B–100009), (V) *O. unisulcatus* (DFM-FRA50t), (W) indeterminate *Otomys* sp. (DRS, Bag 8), (X) Otomys sp. (EPF–Bulk-12).

The molar F (from Sterkfontein Cave) shows similarities with *O. karoensis*, although it was the closest match to *O. sloggetti* out of a large number of M_1 molars examined from the Sterkfontein Name Chamber deposits. This lack of an obvious match to the M^3 *O.* cf. *sloggetti* at Sterkfontein could be due to a number of circumstances; however, it is worth considering the possibility that the M^3 molars, which admittedly look convincingly like *O. sloggetti*, belong to a taxon which has an M_1 which differs from that of *O. sloggetti*. The fossils G, J, K, L, M, N, O, P, Q and R were all originally identified as *O.* cf. *sloggetti,* but when they are lined up and compared it becomes clear that there is a good deal of variation, and, as noted in the discussion above, we currently have no understanding of the link between the different morphotypes and if they represent the same taxon. The fossils J, K, L, M, N and O (also possibly 'P' although digestion and wear complicate assessment) all appear to represent the same morphotype and come from KRM, PP13B, PP30, and KEH1. This morphotype (called Morphotype 3) is represented by an anterior lamina which is rectangular in shape and embedded at an angle into the second laminae (size range: 2.8–3.3 mm length, 1.7–1.9 mm breadth). The modern *Otomys* sp. from Diepkloof Rock Shelter (Q) and the specimen from KDS (R) have more rounded first laminae, and the orientation of the laminae differs from the other fossils, and from each other. Both these specimens fall out of the size ranges shown by Morphotype 3, and are smaller in both length and breadth. The modern Diepkloof specimen (Q; 2.7 mm length, 1.5 mm breadth) shows a strong morphological resemblance to the fossil *Otomys* species from PP30 (G; 2.8 mm length, 1.7 mm breadth) although it is smaller. Differentiating morphotypes in the M_1 is thus less straightforward than seen in the M^3, and although specimens J, K, L, M, N, O and P appear to represent one morphotype (referred to as Morphotype 3 from this point), there are a number of other specimens, such as G and Q (Morphotype 4), and R (which is relatively narrow, and unique in having a small round anterior lamina and relatively linearly orientated laminae) which represents another morphotype (Morphotype 5). The linear arrangement of cusps in Morphotype 5 resemble those of the modern specimen *O. sloggetti* (TM–16519), whereas Morphotype 3, Morphotype 4 and Morphotype 5 specimens all share some similarities with the other modern comparative *O. sloggetti* (TM–22669), but are generally smaller. Once again, it is impossible to interpret the observed variation, which could be indicating a different sub-species or species, or may simply be reflecting morphological plasticity in the phenotype. Notably, morphological differences between the fossil specimens and the one modern *O. sloggetti* specimen (TM–22669) are less than that observed between the two modern museum specimens in Figure 7. As observed in the M^3, the two museum specimens differed in the size and distribution of the cusps varying across the occlusal surface, and the extent of curvature of the laminae (see Figure 7, H and I). The two modern *O. sloggetti* (TM–22669, TM–16519) were broader than all the fossil specimens measured (*O. unisulcatus* excluded), although there was some overlap with the fossils in terms of length.

The modern specimen from EPF (EPF–bulk–12, Figure 7, 'X') (hereafter referred to as Morphotype 6) is differentiated from Morphotypes 3, 4 and 5 taxa by a larger (relative to the other laminae) and more rounded anterior cusp – the former have a smaller and more rectangular-rounded shape. This M_1 (two molars with this morphology were recovered from the bulk samples) may match the unidentified *Otomys* maxilla and M^3 from the EPF bulk pellet samples, or it may represent a variation of cf. *O. karoensis*. The similarity observed in the morphology of the M^3 of the EPF specimen with some of the fossils (i.e. Morphotype 1), is not as obvious when it comes to the M_1, and it is not possible to match the EPF M_1 and M^3 specimens with any certainty.

Discussion and conclusions

An intrasite examination of fossil specimens identified as *O.* cf. *sloggetti* during previous research indicates that, although there is homogeneity observed between some specimens, there is a degree of variation, which, together with the differences observed between modern comparative *O. sloggetti* material, is very hard to interpret. There are, however, two morphotypes distinguishable in the fossil *O.* cf. *sloggetti* M^3, and four in the M_1, material, as described above. We suggest that the differentiation of these morphotypes, at least in the case of Morphotypes

1 and 2 in the case of the M^3, and Morphotype 3 in the case of the M_1, is validated by the fact that several specimens of the same morphotypes have been found in several south coast fossil sites, a couple of west coast sites, and in both fossil and modern material. In addition, the morphotypes were recognised as differing morphologically to *O. karoensis* and *O. irroratus*, and most closely resembling *O. sloggetti*, by two independent researchers. This, together with the fact that the fossil record indicates the presence of the morphotypes from ~150 000 ka, suggests that they have some kind of phenotypic origin and are not merely local aberrations. It is not, however, possible to match the upper and lower morphotypes with any certainty, which is indicative of the fact that we have no clear understanding of the relevance of the different morphotypes, and their genetic affiliations. It is also not possible to ascertain the relationship between the different morphotypes with *O.* cf. *sloggetti*, or other *Otomys* taxa. There is a good morphological match between the unidentified *Otomys* from the EPF bulk sample, and the fossil Morphotype 1, but at this stage it is impossible to say if they represent the same, or some closely related, taxon. The precise relationship of the EPF bulk sample *Otomys* taxon to the fossil (and modern) taxa remains uncertain given all the unknown variables, and the evidence for appreciable genetic divergence in O*. sloggetti*, and other Otomyini. In conclusion, both the fossil and modern specimens suggest a large degree of variation in molar morphology and size, and both Morphotype 1 and Morphotype 2 in the case of the M^3, and Morphotypes 3, 4 and 5 in regard to the M_1, show affinities with *O. sloggetti*. The genetic relevance of the observed similarities and differences will remain obscure until they can be interpreted within a framework which clarifies the link between skull and molar morphology, and the genotype of modern taxa.

The presence of *O.* cf. *sloggetti* in surface deposits in some of the Pinnacle Point sites was puzzling, and it was suggested that this might have occurred due to mixing with fossil deposits.[7] This review of the purported *O. sloggetti* material from the owl pellets and fossil sites suggests that it is probably not *O. sloggetti*, and, as noted above, the jury remains out as to the exact phylogeny of the different morphotypes. The observed differences between the two *O. sloggetti* museum specimens illustrate the fact that museum collections may not provide a reliable taxonomic reference where cryptic species are involved, and are likely to be related to the fact that the specimens were collected from different mountain ranges, and probably represent different species.[2]

It has been noted that specimens identified as *O. sloggetti* in the field were found (in an investigation utilising nuclear and mitochondrial gene regions, as well as morphological characteristics[8]) to represent two clades. One was consistent with the description of *O. sloggetti,* but the other was distinct not only from *O. sloggetti*, but from all other *Otomys* species, and the authors concluded that this could be a new, and novel, species. Subsequent to this, a new *Otomys* species was described from the Sneeuberg Centre of Floristic Endemism in the southern Drakensberg Mountains (*Otomys willani* sp. nov.), and the same publication notes the presence of two specimens from the Mountain Zebra National Park in the Eastern Cape which showed a well-supported and relatively deeply divergent lineage that has no supported relationship with any other Otomyini species.[2] The fossil and modern material discussed in this paper suggests that such cryptic, undescribed species, or subspecies, have been, and may be, co-occurring with modern populations of *O. karoensis* and *O. irroratus*. Phylogenetic studies need to be done in conjunction with morphological studies of skulls and molars, as, currently, the relationship between the huge variation seen in interspecific morphology with genetics is little understood, different *Otomys* species are not always distinguishable morphologically, and considerable chromosomal polytypes have been found.[2] The otomyines are an interesting family of murids and assessing the degree of morphological variation in modern *O. sloggetti, O. irroratus*/*O. auratus,* and *O. karoensis* populations will provide essential information on morphological variation, and how it relates to biogeography and adaptation, and will in turn provide interesting ecological and evolutionary information on a family of small mammals which exhibit great adaptability and phenotypic plasticity. Clearly, further genetic and cladistic research is needed to unravel the cryptic and complicated taxonomy of *Otomys* taxa. Given the issues raised by this research into the south and west coast fossil record, we recommend that future research into the Otomyini focus on modern collections (including owl pellet collections) from the south and west coasts and the Eastern Cape, using an integrated molecular, karyotypic and morphological approach.

Acknowledgements

The support of the DST–NRF Centre of Excellence in Palaeosciences (to T.M.) and SFF Centre for Early Sapiens Behaviour (to T.H.N.) towards this research is hereby acknowledged. Financial support was also received from the National Research Foundation of South Africa (T.M.) and the National Research Council of Norway (T.H.N.), and is duly acknowledged. Thanks are due to Curtis Marean (SACP4: South Africa Coast Paleoclimate, Paleoenvironment, Paleoecology, Paleoanthropology Project) and Naomi Cleghorn who led the Pinnacle Point and Knysna Head excavations, respectively; Christopher Henshilwood who led the Klipdrift Shelter excavations; and Sarah Wurz who led the Klasies River excavations. We also thank Prof. John Parkington and Louisa Hutton (University of Cape Town) who provided access to the Dunefield Midden and Diepkloof Rock Shelter micromammal material, and to Margaret Avery, Cape Nature staff and Ole Fredrik Unhammer for help in locating and collecting modern owl pellets from the De Hoop Nature Reserve. Kathleen Kuman and Bernhard Zipfel kindly provided access to the Sterkfontein fossil material. A big thank you to Muofhe Tshibalanganda, Stefan le Roux and Anton du Plessis at the Central Analytical Facilities CT scan facility (Stellenbosch University) for all their assistance and support during the scanning and rendering of fossil and modern material.

Competing interests

We declare that there are no competing interests.

Authors' contributions

T.M. was the project leader and carried out all the CT scanning on extant and fossil material, and took the images for the publication with the exception of Figure 2 which was done by T.H.N. T.M. and T.H.N. both contributed to the various drafts leading up to the submission of this article. Both T.M. and T.H.N. provided fossil material, and assessed the morphotypes reported on in this paper.

References

1. Monadjem A, Taylor PJ, Denys C, Cotterill FPD. Rodents of sub-Saharan Africa: A biogeographic and taxonomic synthesis. Berlin/Munich/Boston: Walter de Gruyter; 2015.

2. Taylor PJ, Kearney T, Dalton DL, Chakona G, Kelly CM, Barker NP. Biomes, geology and past climate drive speciation of laminate-toothed rats on South African mountains (Murinae: *Otomys*). Zool J Linn Soc. 2019;189(3):1046–1066. https://doi.org/10.1093/zoolinnean/zlz134

3. Taylor PJ, Maree S, Van Sandwyk J, Baxter R, Rambau RV. When is a species not a species? Uncoupled phenotypic, karyotypic and genotypic divergence in two species of South African laminate-toothed rats (Murinae: Otomyini). J Zool. 2009;277(4):317–332. https://doi.org/10.1111/j.1469-7998.2008.00542.x

4. Taylor PJ, Lavrenchenko LA, Carleton MD, Verheyen E, Bennett NC, Oosthuizen CJ, et al. Specific limits and emerging diversity patterns in east African populations of laminate-toothed rats, genus *Otomys* (Muridae: Murinae: Otomyini): Revision of the *Otomys typus* complex. Zootaxa. 2011;3024:1–66. http://hdl.handle.net/2263/18125

5. Engelbrecht A, Taylor PJ, Daniels SR, Rambau RV. Cryptic speciation in the southern African vlei rat *Otomys irroratus* complex: Evidence derived from mitochondrial *cyt b* and niche modelling. Biol J Linn Soc. 2011;104(1):192–206. https://doi.org/10.1111/j.1095-8312.2011.01696.x

6. Edwards S, Claude J, Van Vuuren BJ, Matthee CA. Evolutionary history of the Karoo bush rat, *Myotomys unisulcatus* (Rodentia: Muridae): Disconcordance between morphology and genetics. Biol J Linn Soc. 2011;102(3):510–526. https://doi.org/10.1111/j.1095-8312.2010.01583.x

7. Matthews T, Marean CW, Cleghorn N. Past and present distributions and community evolution of Muridae and Soricidae from MIS 9 to MIS 1 on the edge of the Palaeo-Agulhas Plain (south coast, South Africa). Quat Sci Rev. 2019;235, Art. #105774. https://doi.org/10.1016/j.quascirev.2019.05.026

8. Phukuntsi MA, Kearney T, Brettschneider H, Dalton DL, Oosthuizen M. Hidden Identities: Cryptic species in the *Otomys* genus (Cuvier 1824) (Rodentia: Muridae: Otomyinae) revealed by mitochondrial and nuclear DNA in South Africa. J Phylogenet Evol Biol. 2016;4, Art. #168. http://doi.org/10.4172/2329–9002.1000168

9. Taylor PJ, Denys C, Mukerjee M. Phylogeny of the African murid tribe Otomyini (Rodentia), based on morphological and allozyme evidence. Zool Script. 2004;33:389–402. https://doi.org/10.1111/j.0300-3256.2004.00157.x

10. Denys C. Phylogenetic affinities of the oldest East African *Otomys* (Rodentia, Mammalia) from Olduvai Bed I (Pleistocene, Tanzania). Neues Jahrb Geol Palaontol Abh. 1989;12:705–725. https://doi.org/10.1127/njgpm/1989/1990/705

11. Chevret P, Denys C, Jaeger JJ, Michaux J, Catzeflis F, Molecular and palaeontological aspects of the tempo and mode of evolution in *Otomys* (Otomyinae: Muridae: Mammalia). Biochem Syst Ecol. 1993;21:123–131. https://doi.org/10.1016/0305-1978(93)90017-I

12. Wilson DE, Reeder DM, editors. Superfamily Muroidea. Mammal species of the world: A taxonomic and geographic reference. Baltimore, MD: JHU Press; 2005. p. 894–1531.

13. Avery DM. A preliminary assessment of the micromammalian remains from Gladysvale Cave, South Africa. Palaeontol Afr. 1995;32:1–10.

14. Avery DM. Late Quartenary incidence of some micromammalian species in Natal. Durban Museum Novitates. 1991;16(1):1–11.

15. Avery DM. Notes on the systematics of micromammals from Sterkfontein, Gauteng, South Africa. Palaeontol Afr. 2000;36:83–93. http://hdl.handle.net/10539/16381

16. Denys C. Of mice and men. Evolution in East and South Africa during Plio-Pleistocene times. African biogeography, climate change and human evolution. New York: Oxford University Press; 1999. p. 226–252.

17. Matthews T, Marean C, Nilssen P. Micromammals from the Middle Stone Age (92–167 ka) at Cave PP13B, Pinnacle Point, south coast, South Africa. Palaeontol Afr. 2009;44:112–120.

18. Nel TH, Henshilwood CS. The small mammal sequence from the c. 76–72 ka Still Bay levels at Blombos Cave, South Africa – Taphonomic and palaeoecological implications for human behaviour. PLoS ONE. 2016;11(8), e0159817. https://doi.org/10.1371/journal.pone.0159817

19. Nel TH, Wurz S, Henshilwood CS. Small mammals from Marine Isotope Stage 5 at Klasies River, South Africa – Reconstructing the local palaeoenvironment. Quat Int. 2018;471:6–20. https://doi.org/10.1016/j.quaint.2017.08.074

20. Pillay N, Taylor P, Baxter R, Child MF. *Otomys sloggetti*. The IUCN Red List of Threatened Species [webpage on the Internet]. c2017 [cited 2019 Sep 09]. e.T15659A110019838. http://dx.doi.org/10.2305/IUCN.UK.2017-2.RLTS.T15659A110019838.en

21. Schwaibold U, Pillay N. Behavioral strategies of the African ice rat *Otomys sloggetti robertsi* in the cold. Physiol Behav. 2006;88(4–5):567–574. https://doi.org/10.1016/j.physbeh.2006.05.013

22. Mokotjomela T, Schwaibold U, Pillay N. Population surveys of the ice rat *Otomys sloggetti robertsi* in the Lesotho Drakensberg. Afr Zool. 2010;45(2):225–232. https://doi.org/10.1016/j.physbeh.2006.05.013

23. Taylor PJ, Denys C, Mukerjee M. Phylogeny of the African murid tribe Otomyini (Rodentia), based on morphological and allozyme evidence. Zool Script. 2004;33:389–402. https://doi.org/10.1111/j.0300-3256.2004.00157.x

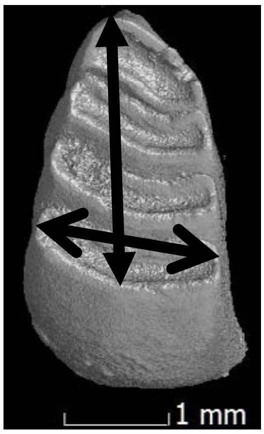
24. Henshilwood CS, Van Niekerk KL, Wurz S, Delagnes A, Armitage SJ, Rifkin RF, et al. Klipdrift shelter, southern Cape, South Africa: Preliminary report on the Howiesons Poort layers. J Archaeol Sci. 2014;45:284–303. https://doi.org/10.1016/j.jas.2014.01.033

25. Matthews T, Denys C, Parkington JE. The palaeoecology of the micromammals from the late middle Pleistocene site of Hoedjiespunt 1 (Cape Province, South Africa). J Hum Evol. 2005;49:432–451. https://doi.org/10.1016/j.jhevol.2005.05.006

26. Avery DM, Stratford DJ, Sénégas F. Micromammals and the formation of the Name Chamber at Sterkfontein, South Africa. Geobios. 2010;43(4):379–387. https://doi.org/10.1016/j.geobios.2010.01.002

Appendix 1: Length and breadth measurements of molars illustrated in Figure 5 (M^3) and Figure 7 (M_1)

	M_1			M^3	
	Length (mm)	**Breadth (mm)**		**Length (mm)**	**Breadth (mm)**
A	3.15	1.94	**A**	2.61	1.43
B	2.15	1.42	**B**	3.03	1.74
C	2.65	1.80	**C**	3.04	1.74
D	3.05	1.69	**D**	3.57	1.91
E	2.87	1.71	**E**	3.58	1.86
F	2.57	1.71	**F**	3.34	N/A
G	2.84	1.74	**G**	3.21	1.68
H	3.48	2.23	**H**	3.50	1.81
I	3.51	2.21	**I**	2.93	1.78
J	3.27	1.89	**J**	2.89	1.80
K	3.06	1.77	**K**	2.52	1.50
L	3.08	~1.84	**L**	2.79	1.81
M	2.86	1.69	**M**	2.44	1.47
N	3.13	1.84	**N**	2.52	1.66
O	3.15	1.74	**O**	3.52	2.17
P	2.83	1.73	**P**	3.60	2.07
Q	2.67	1.51	**Q**	3.24	~2.14
R	2.69	1.48	**R**	3.81	2.22
S	2.90	1.81	**S**	~3.66	2.03
T	3.64	2.14	**T**	~4.07	2.02
U	3.50	2.00			
V	3.82	2.14			
W	3.15	1.98			



Check for updates

AUTHORS:
Robert Luetkemeier[1,2]
Meed Mbidzo[3]
Stefan Liehr[1,2]

AFFILIATIONS:
[1]Institute for Social-Ecological Research, Frankfurt, Germany
[2]Senckenberg Biodiversity and Climate Research Centre, Frankfurt, Germany
[3]Department of Agriculture and Natural Resources Sciences, Namibia University of Science and Technology, Windhoek, Namibia

CORRESPONDENCE TO:
Robert Luetkemeier

EMAIL:
luetkemeier@isoe.de

DATES:
Received: 03 Jan. 2020
Revised: 10 July 2020
Accepted: 27 July 2020
Published: 29 Jan. 2021

HOW TO CITE:
Luetkemeier R, Mbidzo M, Liehr S. Water security and rangeland sustainability: Transdisciplinary research insights from Namibian–German collaborations. S Afr J Sci. 2021;117(1/2), Art. #7773. https://doi.org/10.17159/sajs.2021/7773

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:
interdisciplinarity, transdisciplinarity, sustainable development, co-knowledge production, Namibia

FUNDING:
German Federal Ministry of Education and Research

Water security and rangeland sustainability: Transdisciplinary research insights from Namibian–German collaborations

The Global South is facing severe challenges in ensuring livelihood security due to climate change impacts, environmental degradation and population growth as well as changing lifestyles. These complex problems cannot be solely solved by single scientific disciplines – they require transdisciplinary research (TDR). Stakeholders from civil society, the corporate sector, government and science need to pool their knowledge to find solutions for sustainable transformations. In Namibia, we have been involved in TDR projects on water supply, and sanitation services as well as livestock management in rangeland systems. In this paper, we review two TDR projects that differ in multiple ways and hence allow us to carve out structural differences and critically discuss research outcomes, lessons learned and the challenge of North–South collaborations. Our review builds upon published and unpublished project documents as well as expert interviews with Namibian and German researchers who were involved in the projects. Our results show that TDR can be put into practice in different ways, depending on the research focus and the period available. The TDR phases of problem framing, inter- and transdisciplinary integration were implemented with different tools and foci points. We discuss the role of project length and funding conditions for project success and outcome generation. In addition, we critically consider the role of Namibian and German researchers in these international collaborations. The conclusions we draw touch upon the points of preparatory research funding, the equal acknowledgement of Global South contributions to joint research projects and the explicit handling of TDR components in project work.

Significance:

- The current social-ecological challenges are complex and require TDR as a mode of knowledge co-production, particularly in a development context.
- Inter- and transdisciplinary integration are critical processes for a project to be successful and require the allocation of adequate time and monetary resources.
- Longer-term projects with a funded preparatory research phase constitute a structural model for TDR as project outcomes can evolve over time.
- Global South researchers carry a hidden burden in international collaborations that has to be adequately acknowledged upfront in project planning and final products.

Introduction

The challenges humanity faces today, especially in countries south of the Sahara, are unprecedented. Poverty, inequality and hunger as well as unsafe water supply, unimproved sanitation and infectious diseases remain some of the key problems for millions of people[1], although progress is visible with more and more people reaching good standards of living[2]. These societal challenges are interrelated with climate change impacts, environmental degradation, population growth, urbanisation and changing lifestyles in a complex pattern.[3] They become 'wicked problems' for which by definition 'one-size-fits-all' solutions are not available.[4]

The Sustainable Development Goals (SDG) create a positive vision of the future by targeting good living conditions for all by 2030.[5] However, the key question remains: how can these conditions be created against the background of complex non-linear cause–effect relations in social-ecological systems? For decades, global-scale approaches such as the World3 model[6], the Planetary Boundaries[7] and the 'Doughnut'[8] attempted to disentangle social-ecological interactions, claiming to provide applicable knowledge for societal problems. However, apart from their ability to guide global agenda setting and to raise awareness of humanity's role in the Anthropocene[9], their power to inform local decision-making in order to guide sustainable transformations[10] remains to be tested[11].

The local scale is, however, the most relevant level at which sustainable transformations take place.[12] With increasing interest, transdisciplinary research (TDR) is regarded by practitioners and scholars as a suitable approach to design sustainable development paths adapted to local particularities.[13] Challenges in natural resources management are complex in ontological, analytical and social terms and require decisive boundary work for transdisciplinary teams to achieve outcomes that can be taken up by practitioners.[14] Therefore, TDR is not considered a theory, a methodology or an institution, but rather an approach to developing solutions for societal challenges in which relevant and applicable knowledge is co-produced by both scientists and stakeholders.[15] Particularly against the background of sub-optimal outcomes from non-integrated research and development projects, e.g. in the form of persisting implementation gaps[16], TDR can be considered a promising approach to solve sustainability challenges, especially in the Global South[17]. For this solution to be fruitful, facilitating factors such as the importance of project co-creation and the role of knowledge brokers against issues such as preconceived assumptions, silo-thinking and terminological diversity have already been identified.[18] Despite the growing recognition of TDR and its potential, particularly in sub-Saharan Africa, capacities of scientists and stakeholders on how to successfully conduct TDR projects are still evolving.[19] Although success criteria of TDR have been identified[20], particular demands persist in both the Global North and South for capacity development to design, implement, monitor and evaluate TDR projects.

In this paper, we, as a Namibian–German team of authors, intend to share our experiences from two TDR projects in Namibia on water and food security[21] and rangeland sustainability[22] to contribute to meet the abovementioned demands. We do not intend to provide a comprehensive evaluation of each project as this would require more resources to acknowledge the diverse set of valuations and expectations from research funders, research providers and research users.[23] Our objectives are rather to (1) present key insights into structural project implementation stages and (2) to discuss the projects' major outcomes and lessons learned as well as the role of German and Namibian researchers within these North–South collaborations.

Material and methods

In order to meet these objectives, we reviewed two research projects that were carried out in Namibia between 2006 and 2017. The following sub-sections will briefly shed light on the case study approach we followed, the expert interviews we conducted and the TDR process we used as a reference to structure our results.

Case study approach

TDR projects provide a context-specific forum for all parties involved to learn from and with each other. This mutual learning process enables them to 'learn to collaborate while collaborating' as Freeth and Caniglia[24] aptly phrased it recently. While most project parties may benefit from this, it remains an asset that is hard to access for third parties. Therefore, we consider a case study approach as a suitable tool to provide third-party scientists and practitioners with insights from past TDR projects.

Taking a closer look at the scientific TDR literature confirms this impression as case study reports are a common format to make TDR insights available to third parties. For instance, Roux et al.[23] report on research programmes from South Africa on water management, land-use change, food production and river system health and propose a co-reflection framework that facilitates the evaluation of TDR projects. Cundill et al.[25] reviewed projects from the USA and South Africa on urban water management and freshwater conservation to explore the role of 'communities of practice' in facilitating knowledge sharing and mutual learning processes. Freeth and Caniglia[24] share their experiences from a project that took a systemic perspective on leverage points to identify root causes of unsustainability.

Here, we take a closer look at two projects that were conducted as joint collaborations among Namibian and German partners. We deliberately chose two projects that differ from each other in multiple ways, i.e. in their thematic focus, the time periods available and the degree of stakeholder involvement. We report on the CuveWaters[21] project in which an Integrated Water Resources Management (IWRM) scheme was developed for northern Namibia and on the project OPTIMASS[22] in which rangeland sustainability was assessed on freehold farms in the northeast of Namibia.

Expert interviews

While a comprehensive evaluation of the two projects is not the subject of this paper, we intend to shed light on particular issues of TDR that are rather intangible and often not made explicit during and after project work. In this regard, the term 'outcomes' refers to long-term changes in social-ecological structures and processes (e.g. new routines, governance structures) that go beyond short-term project outputs such as reports, papers and toolkits.[26] We also critically discuss lessons learned from both projects with regard to the way TDR was initiated and carried out. Against these experiences, we try to provide constructive suggestions on how to bypass respective challenges in future projects. Finally, we reflect upon the roles of Namibian and German researchers in these joint collaborations that were funded by the German Federal Ministry for Education and Research. Therewith, we want to provide a critical contribution to the design and implementation of TDR in North–South collaborations.

In order to achieve these objectives, a review of published and unpublished internal reports is a necessary first step. However, we assumed that more detailed information could be obtained from former project members who were involved in the design, implementation and finalisation phases of the projects. Semi-structured, systematising expert interviews[27] were hence regarded as a suitable tool to further expand our own experiences and the information we gained from the literature. Expert interviews are common tools to obtain information, perceptions and viewpoints on structural aspects of TDR projects. For instance, Zscheischler et al.[20] carried out qualitative expert interviews on success factors of TDR to compile a quantitative questionnaire thereof. Defila and Di Giulio[28] consider the challenge of knowledge integration and carried out qualitative interviews with researchers and stakeholders of multiple projects from a common funding scheme on project organisation, implementation, collaboration, common goals and the design of synthesis products.

For the current study, we carried out five systematising expert interviews with former project members – three Namibians and two Germans – who were leading researchers on the projects. The semi-structured interview guideline consisted of open questions on (1) major project outcomes, (2) shortcomings and lessons learned as well as (3) the role of Namibian and German researchers. The interviews provided room for additional comments and statements, while three of them were conducted as face-to-face conversations and two experts provided their answers in written form. All interviews were noted, digitised and coded for a qualitative content analysis. The coding scheme was deducted from the interview guideline and inductively adjusted while working on the transcripts. Ethical clearance to conduct the interviews was provided by the Ethics Commission of the Institute for Social-Ecological Research, Frankfurt, Germany.

Transdisciplinary research process

Today's wicked social-ecological problems require a mode of science that can guide sustainable transformations. While notions of this mode of research may vary from 'mode 2 science', 'action research', 'team science' or 'post-normal science', the conceptual target behind these notions is to generate 'socially robust' knowledge.[15] This understanding acknowledges the necessity and strengths of mono-, multi- and interdisciplinary forms of knowledge production[29], but critically highlights the need to go beyond science in order to co-produce new applicable knowledge.

Overall, the term TDR is not strictly defined, resulting in different interpretations and priority settings. To give a brief introduction into the corresponding literature for the African context, research on water conservation, for instance, highlights the role of mutual learning processes among stakeholders and scientists as a key feature of TDR and its role for overall project success.[30,31] Against this background, Nel et al.[32] conceptualised a framework to guide river and wetland conservation planning and co-knowledge production from a South African case study. They propose a three-tiered process of (1) project co-design, (2) knowledge co-production and (3) co-implementation with an iterative process of knowledge sharing and mutual learning among the participating parties.[32] Their perspective comes close to conceptual works on TDR in global research endeavours under the umbrella of the Future Earth programme. In this context, Mauser et al.[13] propose a similar framework with consecutive steps of (1) co-design, (2) co-production and (3) co-dissemination in which different aspects of integration (scientific, international and sectoral) become relevant.

In the projects under consideration here, we applied a conceptual TDR framework that links to the above-mentioned examples.[15] It comprises three idealised phases of (1) problem framing, (2) interdisciplinary integration and (3) transdisciplinary integration as depicted in Figure 1.

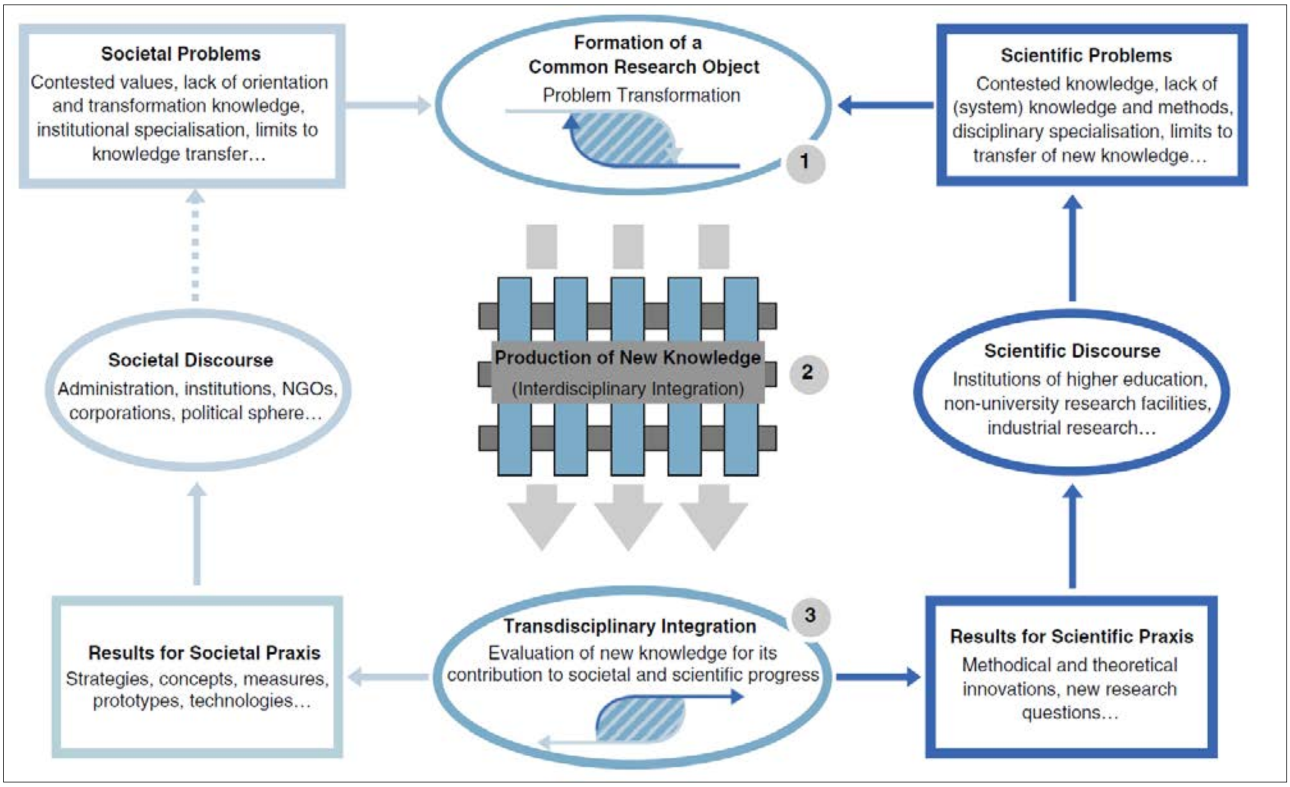


Figure 1: Conceptual model of an idealised transdisciplinary research process.[15]

Initially, phase 1 of problem framing is required in which stakeholders from relevant fields as well as scientists frame their particular interests in order to find a common ground. Ideally, all parties agree upon a 'boundary object' which is basically a shared conceptual understanding of the overall problem context. Once this is found, it can be used to create an 'epistemic object', which is a more precisely defined concept, such as a social-ecological system.[33] This epistemic object serves as a tool for organising knowledge from different stakeholder perspectives and provides the opportunity for mutual learning. From here, phase 2 commences in which a more conventional interdisciplinary science process starts.[15] For this purpose, the overall research question needs to be broken down into smaller pieces that can be worked upon by individual scientific disciplines. In these packages, disciplinary and interdisciplinary work on specific research questions is conducted with two aims: (1) while the major target is to obtain scientific insights that support the overall project goal, (2) the process also targets the production of new knowledge for individual disciplinary communities. The researchers hence work on their specific disciplinary questions that are relevant to their own fields, and need to generate results that are relevant for their colleagues as part of the interdisciplinary effort within the TDR. In phase 3, the results from the individual work packages and disciplines are finally integrated to generate knowledge products that are target group specific and can be fed back into the societal and the scientific discourses.[15]

Results

The results of the project reviews are structured as follows. First, the idealised TDR phases are considered to elaborate upon structural differences between the projects and to shed light on specific aspects of how TDR can be put into practice. Second, major outcomes, lessons learned and challenges in North–South collaborations are presented and critically discussed. At the beginning of each section, the results are summarised in a table while key aspects therein are elaborated on in the associated paragraphs.

Project phases

The idealised TDR process provides a guideline to structure the key aspects of project implementation. We made an attempt to assign certain project activities to each of the TDR phases in order to carve out similarities and differences between the case studies. Table 1 gives a comparative overview of selected project features in this regard.

Problem framing

The initial phase of TDR serves the purpose of finding a common understanding of a given real-world problem from both societal and scientific viewpoints.[15] Theoretically, this process can be considered a bottom-up process in which researchers and stakeholders discuss pressing problems and agree upon shared definitions. In practice however, this requirement is difficult to be met, as the funding frameworks broadly pre-define a thematic area in which researchers and stakeholders are required to operate, although these thematic areas may not necessarily cover the challenges most relevant, locally.

Nevertheless, leaving this strict interpretation aside, both projects managed to facilitate a collaborative problem-framing process. In CuveWaters, the funding agency expected a preparatory phase in which German and Namibian researchers and stakeholders were able to adjust the initial project design according to local needs. This preparatory phase turned out to be a valuable first step to render the project proposal more relevant to local real-world challenges. In the case of OPTIMASS, a preparatory phase was not available, but the project design was still considered efficient by the interviewed experts as German and Namibian researchers could fall back on prior collaborations which facilitated mutual understanding. Nevertheless, full integration of scientists and stakeholders in OPTIMASS was only possible after official project launch. The benefit of preparatory phases or 'quick initiation funding' was already highlighted by scholars like Luthe[34] for TDR research in general and by Giller[35] for North–South collaborations. Respective programmes are now increasingly considered in funding schemes, e.g. by German Federal Ministry for Education and Research and the Dutch Research Council.

Table 1: Comparative overview on key features of the idealised transdisciplinary research stages

Features		CuveWaters	OPTIMASS
Duration		2006 – 2015 (10 years)	2014 – 2017 (3 years)
Focus		Improving water and food security in central-northern Namibia	Shaping sustainable human–nature interactions in rangeland systems of Namibia
Problem framing	Preparation	Overall research scope was pre-defined by funding scheme. Extensive scoping phase for Namibian–German researchers and stakeholders prior to project start (workshops, field visits).	Overall research scope was pre-defined by funding scheme. Prior Namibian–German collaboration facilitated project design. However, start of intensive problem framing after project launch (workshops, field visits).
	Boundary object	'Integrated Water Resources Management' as overall reference concept well known to all partners	'Sustainable Geo-Biosphere Feedback Management' as tangible notion for scientists, stakeholders and farmers
	Epistemic object	'Multi-resources mix' narrative describing the activation of multiple endogenous water potentials for improvement of supply autonomy, food security and job creation	'Ecosystem Services in Rangelands' as depiction of the interconnectedness of livestock-based farm management and rangeland ecosystem responses
Interdisciplinary integration	Major disciplines	Agriculture, geography, sociology, social ecology, hydraulic engineering	Plant ecology, hydrology, geography, social ecology
	Disciplinary work	Integrated work packages on four technologies, each with research questions of (1) technical and institutional design, (2) capacity development for operation and maintenance and (3) social and ecological impact evaluation	Disciplinary work packages on eco-hydrology, vegetation–environment feedbacks, rangeland management and water management
	Integration	The concepts of 'knowledge', 'practices', 'institutions' and 'technology' served as arenas for interdisciplinary exchange within and beyond the work packages	Integration across work packages was established by a formal model that required input from environmental and societal control variables to depict rangeland–ecosystem feedbacks
Transdisciplinary integration	Stakeholders	National and local level institutional stakeholders, primarily governmental/administrative actors were involved over the entire project duration. Ownership of facilities transferred to these stakeholders in final project phase.	Collaboration primarily with local administrative stakeholders and freehold farmers in the study area. Actors from the Namibian agricultural sector were involved for knowledge transfer.
	Population	Technology development and adaptation was co-designed with local population as direct users of new technologies	Farmers were directly engaged for collaborative development of farm-management strategies
	Output	Practice-oriented knowledge products (e.g. manuals, toolkits) and demonstration plants were established to ensure continuation of activities and facilitation of imitators	Knowledge transfer via conference and fair talks, workshops and booklets primarily for farmers

In practical terms, the problem-framing phases in both projects included a number of activities such as inter- and transdisciplinary workshops and field visits. In the case of CuveWaters, the IWRM concept can be considered as the boundary object on which all actors could agree. During the time of project launch, IWRM was the leading paradigm in the water sector[36] to which all stakeholders could connect. The parties involved further concretised what IWRM meant in the project context and created the notion of a multi-resources mix[37] which can be considered the epistemic object of the project. As a narrative, it summarises the broad range of endogenous water resources that should be utilised to sustainably enhance local living conditions.

In a similar manner, researchers and stakeholders in OPTIMASS found a common language by referring to the notion of 'Sustainable Geo-Biosphere Feedback Management' as the weakly structured boundary object. It served as an intuitive representation of how farmers and their livestock interact with the environment. This notion was further concretised by referring to 'Ecosystem Services of Rangelands' – a conceptual understanding which became a key term in current rangeland science in combination with the concept of resilience.[38] This facilitated the integrated discussions on how certain management actions alter ecosystem functions and hence relevant ecosystem services for farmers.

Interdisciplinary integration

The second phase of TDR provides room for new knowledge to be produced by the involved disciplines in a collaborative way. While multiple options to facilitate interdisciplinary integration in a TDR context exist, such as theoretical framing, collaborative formulation of integrated hypothesis or development and application of integrated models[39], a method's suitability for particular projects is always context specific.

Both projects included a range of disciplines from natural, social and engineering sciences, but followed different paths of interdisciplinary integration. CuveWaters adopted an organisational structure in which the different technological interventions were conceptualised as individual work packages. In this regard, each work package was set up in an interdisciplinary way to work on (1) the technical and institutional design, (2) capacity development for operation and maintenance and (3) social and ecological impact evaluation. While interdisciplinary collaboration was hence already conducted at this stage, further arenas for exchange between the work packages were established by regularly discussing progress in empirical and conceptual work under the umbrella concepts of 'institutions', 'practices', 'technologies' and 'knowledge', referred to as the social-ecological structures and processes.[40] This multi-level approach is considered suitable for projects with a larger spectrum of interventions.

OPTIMASS chose a different way by not designing the individual work packages explicitly in an interdisciplinary way but rather integrating the individual disciplinary activities in an overarching formal model that depicts rangeland feedbacks. In this regard, the work packages focused on specific disciplinary tasks in the fields of eco-hydrology, vegetation-environment feedbacks, rangeland management and water management. The project integrated the individual perspectives by using the 'EcoHyD' model which simulates the coupled hydrological and vegetation dynamics taking into account changes in landscape and climate.[41] This was flanked by the conceptual elaboration of the feedback processes between humans and rangelands as a social-ecological system. Although the formal model was not specifically tailored towards the integration of different knowledge domains, the researchers were able to clearly identify their particular contribution to the overall research aim and used the model to put themselves in relation to the other disciplines. Research questions that were addressed in this phase of interdisciplinary integration were (1) which ecological indicators are relevant to farmers in deciding on management actions, (2) which drought adaptation strategies are being carried out, and (3) how these strategies of farmers can be translated in parameters with realistic value ranges for model simulations.

Transdisciplinary integration

Although the third stage of the idealised TDR process may visually be interpreted as the final phase (Figure 1), it can be thought of as a continuous process of mutual learning in which knowledge transfer towards society gains importance by the end of the project.[15] The following paragraphs describe the way institutional stakeholders and individual local actors were involved in the research processes and what types of knowledge products were generated.

CuveWaters involved a broad range of governmental and non-governmental stakeholders that were selected according to their relevance for project implementation and their potential for regional and national policy outreach as well as based on existing personal and institutional relations. In particular, ministerial departments, the national water supplier and river basin committees have been of primary importance and participated regularly in project activities.[42] In terms of local actors, the project particularly involved community members in the design, operation and maintenance of the technological and institutional interventions to ensure their applicability and their post-project continuation. Bottom-up processes were initialised via co-design processes including regular workshops, focus group discussions and trainings.[43] Tailored knowledge products were generated and communicated to the relevant actors. This science–society interface included elements such as workshops, leaflets, films, policy briefs, toolkits and information systems. For instance, a rain and floodwater harvesting toolkit gives practical advice on how to design, construct and operate water harvesting facilities and the 'Interactive Water Information System' provides spatially explicit information on environmental aspects and project results for water managers.[44]

OPTIMASS put a stronger emphasis on national stakeholders from the environmental and agricultural sector. While similar selection criteria were applied as in CuveWaters, prior collaborations facilitated the initial stakeholder engagement process. In this regard, the Ministry of Environment and Tourism and the Namibian farmers' unions were involved. As key landscape managers in the study area, local freehold farmers were engaged in OPTIMASS as the primary addresses and collaborators. Local actors often hold implicit knowledge on ecosystem dynamics that can be valuable to complement and inform explicit scientific knowledge.[45] Local and traditional knowledge systems with a particular focus on natural resources management can also be referred to as 'professional ecological knowledge'.[46] Hence, the farmers were included in the project as equal partners to understand feedbacks in rangeland management and to develop adapted management strategies that build upon currently available practices. While researchers made use of qualitative and quantitative social science methods to obtain first-hand knowledge (consultation of stakeholders[47]), farmers could engage with the project team on an equal basis via workshops, focus group discussions and field experiments (participation of stakeholders[47]).[48] In terms of output, the project designed knowledge products such as booklets with condensed results for farmers and the general public[22] and communicated practical knowledge to the agricultural sector and the farmers beyond the local study area by hosting the '21st Namibian Rangeland Forum' in 2017[49]. Furthermore, mentoring of farmers and involving them in experimental settings was uncovered as a suitable mechanism to transfer knowledge between farmers. In addition, respected experts, e.g. experienced farmers who are acknowledged by their peer community and have a good reputation, can also act as multipliers for scientific knowledge transfer.

Project insights

The following paragraphs will take a closer look at the projects' impacts and what we learned from them. Here, the expert interviews can be considered as the major sources of information to carve out outcomes, lessons learned and challenges in North–South collaborations as presented in Table 2.

Major outcomes

TDR intends to create an impact on social-ecological structures and processes to initiate and accompany transformations towards sustainability.[12] If successful, these long-term impacts can be considered as project outcomes that go beyond direct project outputs such as scientific papers, reports, toolkits and manuals.[26] While Table 2 provides a brief overview of societal and scientific outcomes, the following paragraphs will take a closer look at outcomes on multiple societal levels.

At the local level, CuveWaters intervened with infrastructure and governance systems of water and food supply and hence directly affected the livelihoods of rural and urban inhabitants, mostly disadvantaged people. The consulted experts consider the co-design process to be largely successful, as behavioural routines of inhabitants changed (e.g. new hygienic habits), new occupational opportunities were created (e.g. maintainers/constructors locally known as 'blue team'), local authorities continued facility operations (e.g. sanitation and water reuse system) and households carried on water harvesting and gardening activities (e.g. income generation). As a result of these local level processes, one follow-up project was initiated that builds upon CuveWaters research and further investigates the potential of water reuse for fodder production and thus the alleviation of grazing pressure in northern Namibia (EPoNa).[50] OPTIMASS investigated rangeland feedbacks in livestock-dominated savannah ecosystems and directly collaborated with local freehold farmers. While the three-year project primarily focused on understanding the ecological effects and interactions with societal drivers of rangeland feedbacks, farmers were engaged in developing farm management strategies that were locally applied and tested. The project identified indicators for farm decision-making (e.g. plant species composition and features) and developed recommendations on combining horizontal (e.g. rotational grazing with effects into the area) and vertical measures (e.g. de-bushing with soil-penetrating effects) to enhance soil water and vegetation conditions.[48] This knowledge was locally applied and disseminated to other farmers. As a result of this process, trust between farmers and researchers was created, which led to the initiation of a follow-up project that further investigates degradation processes and tipping points in the same study area (NamTip).[51]

Besides the local level, both projects engaged with regional and national level stakeholders to generate outcomes. CuveWaters collaborated with authorities and non-governmental organisations throughout the country and beyond to disseminate the knowledge gained. In this regard, multiple spin-off initiatives were triggered that adopted the sanitation and water reuse approach as well as the rain and floodwater harvesting technologies for new problem contexts. Furthermore, an attempt was made to introduce the dualistic vocational training approach that proved to be successful in the project to the national regulations of formal training and education. OPTIMASS was able to feed the results into the national discourse, primarily by co-organising a national conference of the agricultural sector and by engaging with national-level stakeholders. In addition, a spin-off project could be initiated in a different region that investigates wildlife-based strategies as one option for sustainable rangeland management (ORYCS).[52]

Table 2: Overview on selected outcomes, lessons learned and issues of North–South collaboration

Features		CuveWaters	OPTIMASS
Project outcomes	Society	Co-developed interventions led to new behavioural routines (hygiene practices) and occupational opportunities (income from gardening) Knowledge on sanitation and water reuse technologies disseminated to other cities Low-tech rain and floodwater harvesting and gardening techniques were adopted by NGOs/households and implemented elsewhere	Freehold farmers gained knowledge on ecosystem responses of certain livestock farming practices and about options for adaptation strategies for rangeland management Trust among participating farmers and researchers was created as willingness to collaborate in and openness towards follow-up research projects increased
	Science	Insights into community organisation for efficient operation of rainwater and gardening facilities 'Demand-responsive' assessment procedure was developed to facilitate co-design of technologies and institutions Intensified and sustained collaboration between researchers and institutions with follow-up projects	Follow-up projects initiated that continue ecosystem degradation research (NamTip) and wildlife management challenges (ORYCS) New research questions on wildlife–livestock interactions as well as bioturbation were developed Intensified and sustained collaboration between researchers and institutions
Lessons learned	Insights	Long project duration facilitated deep transdisciplinary research integration as stakeholders could be engaged over a long period Local institutional stakeholders show more ownership than national actors Co-design approach successful as most facilities are operated and maintained 'Scientific empathy' more important than knowledge on TDR methodology	Engagement of local farmers worked via hands-on demonstrations and field experiments Farmers' participation required scientists to partly rethink their initial research focus and approach Longstanding contacts to target community are valuable catalysts for trust building in short-term projects
	Shortcomings	High turn-over of decision-makers required iterative trust building Knowledge transfer on national level was bigger challenge than in the local context Pre-defined research scope by donor constrained a full bottom-up process Unclear responsibilities of institutional stakeholders for continuation of some activities	Short project duration prevented a deep stakeholder integration Project output was more scientific (e.g. papers, concepts, models) and less targeted towards direct practical benefits to farmers Only the initiated follow-up projects served to create more tangible practice-oriented output
North–South	Collaboration	Namibian researchers facilitated the institutional, organisational and cultural conduction of empirical research German researchers provided the technical/engineering knowledge Research team showed mutual respect and acknowledged individual strengths	Project lead was shared among German and Namibian partners Namibian researchers engaged in stakeholder consultation, logistics, student supervision and ecological assessments German researchers focused primarily on modelling and ecological assessments
	Challenges	Project coordination was assigned to the German partners due to funding conditions, with challenges for local integration and communication Balancing scientific with political interests on national and local levels regarding project locations and research questions Adaptation of 'German knowledge system' to carry out empirical research in Namibia	Project lead was shared but perception of German dominance persisted In-kind contributions of the South were not acknowledged appropriately compared to monetary funds from German donors Namibian researchers had 'double role' as local (cultural, organisational) facilitators and researchers, which led to high workload and potential trade-offs

Overall, both projects managed to generate impacts on social-ecological structures and processes with different intensities. One major difference between them is the period available and the research focus. While CuveWaters could engage with the local population over a long period and explicitly target technological and institutional interventions, OPTIMASS rather focused on understanding farmer–rangeland interactions. As a result, tangible societal project outcomes are more obvious from the CuveWaters than from the OPTIMASS project.

Lessons learned

The previous section described the impacts of the projects on multiple societal levels. Now, to critically reflect upon both projects, our expert interviews provide some insights into aspects of TDR that can be considered successful and others that fell short of their expectations.

First, it turns out that project duration is a critical factor to meet TDR requirements. CuveWaters was able to engage with local stakeholders over a period of 10 years while the OPTIMASS project only had 3 years available. In particular, projects that strongly interfere with basic infrastructural settings of a region, such as adjusting the water and food supply systems, require a longer time horizon to embed project activities into local institutional settings as much as possible. Projects with shorter periods do not have the time and monetary means available to do so. For TDR projects with typical periods of about 3 years, the consulted experts identified the need for funding agencies to rethink funding schemes as sustainable project impacts are likely to require longer time horizons to evolve. These longer time horizons may be achieved either via longer funding periods or more options for follow-up projects that continue existing collaborations.[34] However, with longer periods, other challenges appear as, for instance, policy- or decision-makers may change over time, requiring new trust-building incentives from both sides.

Second, following a co-design approach to develop technological and institutional interventions as well as management strategies proved to be successful. CuveWaters developed a 'demand-responsive' approach that serves to design locally adapted institutional set-ups in a mutual learning process between researchers and inhabitants.[53] In the case of OPTIMASS, the co-design process even required a partial shift of the initial research focus on water management as this was not the main issue to farmers. Their problem perception led to surprises in the research results and at the same time changed the perspective of the analysis. Furthermore, it proved to be beneficial for knowledge transfer to provide hands-on examples and demonstrations of particular solutions on the ground so that inhabitants and farmers could clearly grasp the intended benefits.

Third, the experts interviewed pointed out that openness of involved researchers and stakeholders towards one another are critical success factors for TDR projects. This kind of 'scientific empathy' might even be more important than well-founded knowledge in TDR methods[39], although certain actors require this set of knowledge to guide the process. Hence, further training in TDR methodology is required, but more emphasis may be given towards the establishment and reproduction of an open-minded atmosphere of mutual respect and acknowledgement. This insight may be of particular importance for the 'undisciplinary journey' of early-career researchers who seek to work in a TDR environment with specific challenges, as recently discussed by Haider et al.[54].

North–South collaboration

International collaborations between partners from the Global North and Global South in the field of sustainability science have a long and dynamic history[55,56], with renewed appreciation in the context of the Agenda 2030 process with SDG 17 explicitly targeting strengthened global partnerships[5]. As Saric et al.[55] conclude from an expert survey, mutual trust, joint decision-making and research agenda setting as well as transparency are considered key features for successful international partnerships. However, in practices, these requirements are often not met for research and development projects.[57] Here, we intend to provide two insights from our case studies on (1) the 'double-role' of Namibian researchers and (2) equal funding conditions and adequate valuation of in-kind contributions.

First, the consulted experts confirm for both projects that Namibian researchers were confronted with a 'double role'. They were not only required to pursue their specific research questions with associated empirical field work and analytical tasks, but also implicitly required to act as local facilitators for the German researchers by guiding them during the field campaigns as cultural mediators. They strongly engaged with national and local institutional stakeholders, facilitated the cultural and practical set-up of socio-empirical surveys and provided supervision of German and Namibian students, especially during the initial project periods. Although the German researchers also engaged in stakeholder processes and student supervision, they rather took over the scientific part of representing the engineering, social science and ecological disciplines. Overall, this may have led to high workloads for the Namibian researchers in both projects, which could have contributed to a perceived lower recognition and visibility of their scientific contributions.

Second, although both Namibian and German researcher teams shared project responsibilities, monetary funds came from the German ministerial side. This fact contributed to the perception that the German researchers were more dominant in the project team in shaping the overall research agenda than were their Namibian counterparts, as they were considered to be in closer contact to the funding agency. In this context, it was highlighted that in-kind contributions from the Namibian partners were not equally valued as the monetary contributions from Germany. These in-kind contributions can, however, add up to significant amounts of working hours and infrastructure provision, specifically considering the necessity to act as local facilitators for German partners. Overall, a more transparent and equal funding mechanism for Namibian researchers and students in line with their German counterparts was considered an important component for future collaborations.

Despite these critical perspectives on the North–South collaboration in the case studies, all experts agreed that mutual trust and the acknowledgement of individual strengths was an important cornerstone in both projects. Nevertheless, the issues of high workload for Global South partners and the fair acknowledgement of in-kind contributions to research and development projects are considered relevant points for successful future TDR projects. These findings are in line with current debates on 'helicopter science' and how North–South collaborations could potentially be founded on a more equal basis.[35]

Conclusions

Current challenges in the Global South can be considered 'wicked' social-ecological problems that require TDR as a mode of science capable of producing 'socially robust' knowledge. Global tools to inform decision-making for sustainable transformations are a good starting point, but local actions are required that take into account relevant particularities. Against this background, we reviewed two TDR projects in Namibia in which challenges in water and food security as well as rangeland sustainability were investigated. Our results provide insights into structural features of TDR projects and their outcomes, lessons learned and challenges in North–South collaborations. With our review as background information, we do not intend to provide new principles or frameworks for TDR implementation. We rather intend to recommend three critical aspects which we think should be considered for successful TDR in the future, for both funders and researchers.

First, we reinforce the recognition that TDR requires revised forms of funding as the mode of research is different from conventional approaches. Achieving positive and long-term project impacts requires more adapted research questions to be formulated that better fit to local real-world problems. In this regard, 'quick initiation funding' or preparatory research phases are promising ways forward. Nevertheless, longer-term funding schemes are required as the initial processes of problem framing and trust building are time-consuming but necessary first steps that provide more benefits the longer collaborations last.

Second, in order to alleviate the power imbalances in terms of monetary funding from the Global North, we recommend acknowledgement of in-kind contributions of Global South partners in a more equitable way and we consider strengthened and independent funding agencies in the Global South as a good way forward (e.g. the Southern African Science Service Centre for Climate Change and Adaptive Land Management). Furthermore, we see the necessity to account for the high workload and double roles of Global South researchers upfront, as this burden is often hidden behind implicit collaboration routines.

Third, while TDR is receiving increased attention, we see the need to make inter- and transdisciplinary activities in a project more explicit to all parties involved. TDR is not just required on paper but needs to be handled explicitly as a critical and continuous activity to which all parties contribute. Hence, we make the case for clear TDR guidance in a project in which responsibilities among project partners are clearly defined and even budgets are allocated to certain activities (e.g. workshops, retreats, co-writing periods). Joint publications with equal contributions from Global North and Global South partners were considered by our interview partners as key elements for interdisciplinary integration that can foster capacity building on both sides. The full potential for knowledge integration and co-production can only be exploited if TDR is considered a 'hard' rather than a 'soft' project component.

Acknowledgements

This paper builds on research experiences from multiple projects throughout the last years. Special thanks go to our colleagues who were involved in the projects CuveWaters, OPTIMASS, NamTip, ORYCS, EPoNA and the Southern African Science Service Centre for Climate Change and Adaptive Land Management. The research was conducted as part of the ORYCS project (01LL1804C) that is funded by the German Federal Ministry of Education and Research (BMBF) in the context of its funding platform 'Research for sustainable development' (FONA).

Competing interests

We declare that there are no competing interests.

Authors' contributions

R.L. conceptualised the paper and wrote the initial draft. R.L., S.L. and M.M. reviewed the projects. R.L. and M.M. interviewed the experts and analysed the interview material. R.L., S.L. and M.M. reviewed the draft.

References

1. UNECA. MDG report: Lessons learned in implementing the MDGs. Addis Ababa: United Nations Economic Commission for Africa; 2015.
2. UN. The Millennium Development Goals report. New York: United Nations; 2015.
3. AUC. Agenda 2063: The Africa we want. Addis Ababa: African Union Commission; 2015.
4. Lazarus RJ. Super wicked problems and climate change: Restraining the present to liberate the future. Environ Law Report. 2010;40(8):749–756.
5. UN. The Sustainable Development Goals report. New York: United Nations; 2016.
6. Meadows DH, Meadows DL, Randers J, Behrens WW (III). The limits to growth: A report for the Club of Rome's project on the predicament of mankind. New York: Universe Books; 1972.
7. Rockström J, Steffen W, Noone K, Persson A, Chapin FS (III), Lambin EF, et al. A safe operating space for humanity. Nature. 2009;461:472–475. https://doi.org/10.1038/461472a
8. Raworth K. A doughnut for the Anthropocene: Humanity's compass in the 21st century. Lancet Planet Health. 2017;1(2), e48-e49. https://doi.org/10.1016/S2542-5196(17)30028-1
9. Galaz V, Biermann F, Folke C, Nilsson M, Olsson P. Global environmental governance and planetary boundaries: An introduction. Ecol Econ. 2012;81:1–3. https://doi.org/10.1016/j.ecolecon.2012.02.023
10. Moser SC. Editorial overview: Transformations and co-design: Co-designing research projects on social transformations to sustainability. Curr Opin Environ Sustain. 2016;20:v–viii. https://doi.org/10.1016/j.cosust.2016.10.001
11. Häyhä T, Lucas PL, Van Vuuren DP, Cornell SE, Hoff H. From planetary boundaries to national fair shares of the global safe operating space – How can the scales be bridged? Glob Environ Change. 2016;40:60–72. https://doi.org/10.1016/j.gloenvcha.2016.06.008
12. Balvanera P, Calderón-Contreras R, Castro AJ, Felipe-Lucia MR, Geijzendorffer IR, Jacobs S, et al. Interconnected place-based social–ecological research can inform global sustainability. Curr Opin Environ Sustain. 2017;29:1–7. https://doi.org/10.1016/j.cosust.2017.09.005
13. Mauser W, Klepper G, Rice M, Schmalzbauer BS, Hackmann H, Leemans R, et al. Transdisciplinary global change research: The co-creation of knowledge for sustainability. Curr Opin Environ Sustain. 2013;5:420–431. https://doi.org/10.1016/j.cosust.2013.07.001
14. Mollinga PP. Boundary work and the complexity of natural resources management. Crop Sci. 2010;50:S-1–S-9. https://doi.org/10.2135/cropsci2009.10.0570
15. Jahn T, Bergmann M, Keil F. Transdisciplinarity: Between mainstreaming and marginalization. Ecol Econ. 2012;79:1–10. https://doi.org/10.1016/j.ecolecon.2012.04.017
16. Gall M, Nguyen KH, Cutter SL. Integrated research on disaster risk: Is it really integrated? Int J Disaster Risk Reduct. 2015;12:255–267. https://doi.org/10.1016/j.ijdrr.2015.01.010
17. Pereira L, Frantzeskaki N, Hebinck A, Charli-Joseph L, Drimie S, Dyer M, et al. Transformative spaces in the making: Key lessons from nine cases in the Global South. Sustain Sci. 2020;15:161–178. https://doi.org/10.1007/s11625-019-00749-x
18. Sitas N, Reyers B, Cundill G, Prozesky HE, Nel JL, Esler KJ. Fostering collaboration for knowledge and action in disaster management in South Africa. Curr Opin Environ Sustain. 2016;19:94–102. https://doi.org/10.1016/j.cosust.2015.12.007
19. Lang DJ, Wiek A, Bergmann M, Stauffacher M, Martens P, Moll P, et al. Transdisciplinary research in sustainability science: Practice, principles, and challenges. Sustain Sci. 2012;7:25–43. https://doi.org/10.1007/s11625-011-0149-x
20. Zscheischler J, Rogga S, Lange A. The success of transdisciplinary research for sustainable land use: Individual perceptions and assessments. Sustain Sci. 2018;13:1061–1074. https://doi.org/10.1007/s11625-018-0556-3
21. Liehr S, Kramm J, Jokisch A, Müller K, editors. Integrated water resources management in water-scarce regions: Water harvesting, groundwater desalination and water reuse in Namibia. London: IWA Publishing; 2018.
22. Blaum N, Lohmann D, Geißler K. A joint Namibian-German research project. University of Potsdam. Potsdam: Plant Ecology & Nature Conservation; 2018.
23. Roux DJ, Stirzaker RJ, Breen CM, Lefroy EC, Cresswell HP. Framework for participative reflection on the accomplishment of transdisciplinary research programs. Environ Sci Policy. 2010;13:733–741. https://doi.org/10.1016/j.envsci.2010.08.002
24. Freeth R, Caniglia G. Learning to collaborate while collaborating: Advancing interdisciplinary sustainability research. Sustain Sci. 2020;15:247–261. https://doi.org/10.1007/s11625-019-00701-z
25. Cundill G, Roux DJ, Parker JN. Nurturing communities of practice for transdisciplinary research. Ecol Soc. 2015;20(2), Art. #22. https://doi.org/10.5751/ES-07580-200222
26. Margoluis R, Stem C, Swaminathan V, Brown M, Johnson A, Placci G, et al. Results chains: A tool for conservation action design, management, and evaluation. Ecol Soc. 2013;18(3), Art. #22. https://doi.org/10.5751/ES-05610-180322
27. Bogner A, Menz W. Das theoriegenerierende Experteninterview: Erkenntnisse, Wissensformen, Interaktion [The theory-generating expert interview: Findings, forms of knowledge, interaction]. In: Bogner A, editor. Das Experteninterview: Theorie, Methode, Anwendung. 2. Aufl. Opladen: Leske + Budrich; 2005. p. 33–70. German.
28. Defila R, Di Giulio A. Integrating knowledge: Challenges raised by the "Inventory of Synthesis". Futures. 2015;65:123–135. https://doi.org/10.1016/j.futures.2014.10.013
29. Stock P, Burton RJF. Defining terms for integrated (multi-inter-trans-disciplinary) sustainability research. Sustainability. 2011;3:1090–1113. https://doi.org/10.3390/su3081090
30. 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;12:711–726. https://doi.org/10.1007/s11625-017-0446-0
31. Reyers B, Roux DJ, Cowling RM, Ginsburg AE, Nel JL, O'Farrell, P. Conservation planning as a transdisciplinary process. Conserv Biol. 2010;24:957–965. https://doi.org/10.1111/j.1523-1739.2010.01497.x
32. Nel JL, Roux DJ, Driver A, Hill L, Maherry AC, Snaddon K, et al. Knowledge co-production and boundary work to promote implementation of conservation plans. Conserv Biol. 2016;30:176–188. https://doi.org/10.1111/cobi.12560
33. Liehr S, Röhrig J, Mehring M, Kluge T. How the social-ecological systems concept can guide transdisciplinary research and implementation: Addressing water challenges in central northern Namibia. Sustainability. 2017;9:1109. https://doi.org/10.3390/su9071109
34. Luthe T. Success in transdisciplinary sustainability research. Sustainability. 2017;9:71. https://doi.org/10.3390/su9010071
35. Giller KE. Grounding the helicopters. Geoderma. 2020;373:114302. https://doi.org/10.1016/j.geoderma.2020.114302
36. Woodhouse P, Muller M. Water governance – An historical perspective on current debates. World Dev. 2017;92:225–241. https://doi.org/10.1016/j.worlddev.2016.11.014

37. Kluge T. Integrated water resources management (IWRM) in water-scarce regions – from theory to implementation. In: Liehr S, Kramm J, Jokisch A, et al., editors. Integrated water resources management in water-scarce regions: Water harvesting, groundwater desalination and water reuse in Namibia. London: IWA Publishing; 2018. p. 1–7.

38. Sala OE, Yahdjian L, Havstad K, Aguiar MR. Rangeland ecosystem services: Nature's supply and humans' demand. In: Briske DD, editor. Rangeland systems. Cham: Springer International Publishing; 2017. p. 467–489.

39. Bergmann M, Jahn T, Knobloch T, Krohn W, Pohl C, Schramm E, et al. Methods for transdisciplinary research. Frankfurt: Campus Verlag; 2012.

40. Mehring M, Bernard B, Hummel D, Liehr S, Lux A. Halting biodiversity loss: How social–ecological biodiversity research makes a difference. Int J Biodivers Sci Ecosyst Serv Manag. 2017;13:172–180. https://doi.org/10.1080/21513732.2017.1289246

41. Tietjen B. Same rainfall amount different vegetation – How environmental conditions and their interactions influence savanna dynamics. Ecol Modell. 2016;326:13–22. https://doi.org/10.1016/j.ecolmodel.2015.06.013

42. Polak M, Kluge T. Water governance in Namibia: Progress in and challenges of putting IWRM into practice. In: Liehr S, Kramm J, Jokisch A, et al., editors. Integrated water resources management in water-scarce regions: Water harvesting, groundwater desalination and water reuse in Namibia. London: IWA Publishing; 2018. p. 23–26.

43. Deffner J, Birzle-Harder B. Social adoption, experiences and constraints of small-scale horticulture with harvested rain- and floodwater. In: Liehr S, Kramm J, Jokisch A, et al., editors. Integrated water resources management in water-scarce regions: Water harvesting, groundwater desalination and water reuse in Namibia. London: IWA Publishing; 2018. p. 111–117.

44. Bischofberger J, Schuldt-Baumgart N, Röhrig J, Schulz O, Neugart M. Knowledge exchange between research and practice. In: Liehr S, Kramm J, Jokisch A, et al., editors. Integrated water resources management in water-scarce regions: Water harvesting, groundwater desalination and water reuse in Namibia. London: IWA Publishing; 2018. p. 225–234.

45. Berkes F. Sacred ecology. 4th ed. London: Routledge; 2018.

46. Fleischman F, Briske DD. Professional ecological knowledge: An unrecognized knowledge domain within natural resource management. Ecol Soc. 2016;21(1), Art. #32. https://doi.org/10.5751/ES-08274-210132

47. Mobjörk M. Consulting versus participatory transdisciplinarity: A refined classification of transdisciplinary research. Futures. 2010;42:866–873. https://doi.org/10.1016/j.futures.2010.03.003

48. Bischofberger J, Reutter C, Liehr S, Schulz O. The integration of stakeholder knowledge – How do Namibian farmers perceive natural resources and their benefits? Paper presented at: Conference on International Research and Food Security; 2016 September 18–21; Vienna, Austria.

49. OPTIMASS. Rangeland in a changing world: The management and understanding of Namibian savanna rangelands. Paper presented at: 21st Namibian National Rangeland Forum; 2017 July 17–19; Otjiwarongo, Namibia.

50. Technical University of Darmstadt. EPoNa Project [webpage on the Internet]. No date [cited 2020 Apr 28]. Available from: http://epona-africa.com/index.php/en/

51. University of Bonn. NamTip Project [webpage on the Internet]. No date [cited 2020 Apr 28]. Available from: https://www.namtip.uni-bonn.de/?set_language=en

52. University of Potsdam. OPTIMASS Project [webpage on the Internet]. No date [cited 2019 Jun 12]. Available from: https://www.optimass.org/index.html

53. Zimmermann M, Jokisch A, Deffner J, Brenda M, Urban W. Stakeholder participation and capacity development during the implementation of rainwater harvesting pilot plants in central northern Namibia. Water Supply. 2012;12:540–548. https://doi.org/10.2166/ws.2012.024

54. 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. 2018;13:191–204. https://doi.org/10.1007/s11625-017-0445-1

55. Saric J, Blaettler D, Bonfoh B, Hostettler S, Jimenez E, Kiteme B, et al. Leveraging research partnerships to achieve the 2030 Agenda Experiences from North-South cooperation. Gaia. 2019;28:143–150. https://doi.org/10.14512/gaia.28.2.13

56. Gaillard JF. North-South research partnership: Is collaboration possible between unequal partners? Knowledge Policy. 1994;7:31–63. https://doi.org/10.1007/BF02692761

57. Bradley M. North-South research partnerships: Challenges, responses and trends: A literature review and annotated bibliography. Ottawa: Spatial Initiatives Division (IDRC); 2007.

Identified main fire hotspots and seasons in Côte d'Ivoire (West Africa) using MODIS fire data

AUTHORS:
Tionhonkélé D. Soro[1]
Moussa Koné[1]
Aya B. N'Dri[1]
Evelyne T. N'Datchoh[2]

AFFILIATIONS:
[1]Laboratory of Ecology and Sustainable Development, Faculty of Natural Sciences, Nangui Abrogoua University, Abidjan, Côte d'Ivoire
[2]Atmospheric Physics Laboratory, Faculty of Structure of Matter and Technology Sciences, Félix Houphouët Boigny University of Cocody, Abidjan, Côte d'Ivoire

CORRESPONDENCE TO:
Tionhonkélé Soro

EMAIL:
soro.drissa1@univ-na.ci

DATES:
Received: 21 Nov. 2019
Revised: 21 June 2020
Accepted: 06 July 2020
Published: 29 Jan. 2021

HOW TO CITE:
Soro TD, Koné M, N'Dri AB, N'Datchoh ET. Identified main fire hotspots and seasons in Côte d'Ivoire (West Africa) using MODIS fire data. S Afr J Sci. 2021;117(1/2), Art. #7659. https://doi.org/10.17159/sajs.2021/7659

ARTICLE INCLUDES:
☒ Peer review
☐ Supplementary material

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

EDITOR:
Bettine van Vuuren

KEYWORDS:
active fires, fire density, fire hotspot, fire season, remote sensing

FUNDING:
Campus France, the Ministries of Europe and Foreign Affairs, and of the Higher Education, Research and Innovation (France) (grant no: mopga-short-0000000184)

Biomass burning has become more frequent and widespread worldwide, with a significant proportion occurring in tropical Africa. Fire dynamics have been generally studied at global or regional scales. At local scale, however, fire impacts can be severe or catastrophic, suggesting local analyses are warranted. This study aimed to characterise the spatio-temporal variations of vegetation fires and identify the main fire hotspots in Côte d'Ivoire, a country of West Africa, one of the world's burn centres. Using MODIS-derived fire data over a 10-year period (2007–2016), the number of fire days, active fires and fire density were assessed across the entire country. In the southern part dominated by forests, fire activity was low. Three main fire hotspots were identified between 2°30'–8°30'W and 7°00'–10°30'N in the North-West, North-East and Central areas all dominated by savannas. In these areas, Bafing, Bounkani and Hambol regions recorded the highest fire activity where fire density was 0.4±0.02, 0.28±0.02 and 0.18±0.01 fires/km²/year, respectively. At national scale, the annual fire period stretched from October to April with 91% of fires occurring between December and February, with a peak in January. Over the decade, there was a decreasing trend of fire activity. Fire density also was negatively correlated with rainfall >1000 mm for the synchronic analysis, whereas fire density was positively correlated with rainfall in the previous years. Results suggest that the positive relationship between the previous year's rainfall and fire activity could operate on a cycle from 1 to 4 years.

Significance:

- Three fire hotspots were found primarily in savanna vegetation, which burns more regularly than forest-dominated vegetation.
- The fire season occurs over 7 months, the majority of active fires (91%) occurring in just 3 months (December-January-February) with a peak in January (39%).
- Fire activity has declined over the past decade with a return time of above-average fires from 1 to 4 years.
- Fire density is positively correlated to the amount of rainfall in preceding years, whereas fire density and rainfall of the same year were negatively correlated in the region of rainfall >1000 mm.

Introduction

Fire is considered to be a major determinant of the distribution and function of the world's savanna types.[1] Fire has played a significant role in shaping these landscapes, particularly in Africa. Indeed, the African continent has been known as 'the fire continent' due to the high number of fire occurrences[2], favoured by precipitation irregularities. These irregularities lead to the alternation of wet and dry seasons, with dry seasons favouring fires.

The ability of humans to manage fires is still imperfect and may become more difficult in the future with climate change and growing human populations which tend to increase the risk of fire ignitions.[3] Although fires have provided several environmental services over millennia, they can also impact negatively on the environment and socio-economic activities, especially in recent decades. For example, the burning of areas that have been deforested due to human actions[4] contributes to increasing greenhouse gases, aerosols and other gases such as carbon monoxide and ozone emission into the atmosphere[5,6]. Thus, vegetation fires have a direct impact on climate change and air pollution. Moreover, frequent fires can cause soil erosion and reduce soil fertility.[7] Most fires that have occurred in tropical ecosystems in the recent decades have impacted human health and have caused billions of dollars of damage to economies[8,9], especially during the dry season in rural areas where fire remains very difficult to control. In addition, uncontrolled fires are frequent in tropical regions and elsewhere, with those of the tropics being less studied.[10] This is the case for West African countries with high fire activity such as Côte d'Ivoire.[11]

Fires in African savannas are modulated by climate that determines burning timing and fire onset. Thereby, global climate change will increase both the risk of extreme fire events[12] and fire frequency[13]. Predicting where and when extreme fire events will occur remains difficult.[14] It is therefore necessary to better understand fire dynamics for future disaster prevention which may allow mitigating fire socio-economic damage. Fire dynamics are more often investigated on global and regional scales. Even if small effects of fires are reported at global scale, their impacts at regional and local scales can be, respectively, severe and catastrophic.[10] In Côte d'Ivoire, fire damage is recurrent and occurs each year in more than half the country. A large percentage of the Ivorian population lives in a fire-prone rural area.[15] Thus, local analysis of temporal and spatial variations of fires in Côte d'Ivoire is necessary in order to identify the fire-prone areas and to better understand fire dynamics, both recent and potential future dynamics. This will provide very useful information for fire users and decision-makers.

To address this subject, the Moderate Resolution Imaging Spectroradiometer (MODIS) data are often used due to the fact that the sensor was specifically designed and developed to detect active fires.[16,17] Several studies have been performed in a range of countries and regions, from meso-scale to local scales with high resolution to investigate the human role and other drivers underlying fire variability.[18,19] For Côte d'Ivoire, some fire studies have focused only on protected areas[20,21], excluding rural areas that are more fire prone. Others have been carried out over several countries[11,22], making it impossible to establish fire dynamics and focus on risk areas at the national level.

The aim of this study was to characterise the spatio-temporal variations of fires and identify the main fire hotspots in Côte d'Ivoire. The study also addressed the following questions:

1. Where are the main fire hotspots in Côte d'Ivoire located?
2. Do the periods of fire occurrence and of fire peaks vary among regions?
3. What are the changes in fire variables over the recent decade?
4. What are the drivers of fire density (fires per unit area)?

Methods

Study area

Location and climate

The study area covered the entire territory of Côte d'Ivoire which extends over 322 462 km^2, located in West Africa between 2°30'–8°30'W and 4°30'–10°30'N (Figure 1). The human population density of the country averages about 80 inhabitants/km^2, varying from one region to another as shown in Table 1.[15] The climate is controlled by the north-south migration of the Inter-Tropical Convergence Zone (ITCZ), similar to the wider West African region. This produces an alternation of wet and dry seasons within each year. Biomass burning patterns also follow the seasonal shift determined by the ITCZ.[23]

Two main climate types can be distinguished:

1. The humid equatorial climate, covering forest areas in the South, is characterised by four seasons with mean temperature of about 28 °C. Two rainy seasons occur: a long one from April to July and a short one from September to November. Two dry seasons also occur: a short one in August and a long one from December to March.
2. The tropical climate in the Central and North is generally characterised by a single rainy season from June to September and a long dry season from October to May. Generally, temperatures vary between 14 °C (mean minimum, in December–January) and 36 °C (mean maximum, in March).

Overall, there is a strongly decreasing precipitation gradient from South to North across the country. The mean annual rainfall varies from 1600 mm/year to 1800 mm/year in the South, in the ombrophilous area; then moving northward, from 1300 mm/year to 1500 mm/year in the semi-deciduous area, from 1100 mm/year to 1300 mm/year in the forest-savanna area, from 800 mm/year to 1000 mm/year in the sub-sudanian area, and from 500 mm/year to 800 mm/year in the northernmost Sudanian area. This south-north precipitation gradient is associated with changes in the dominant vegetation type as described below.

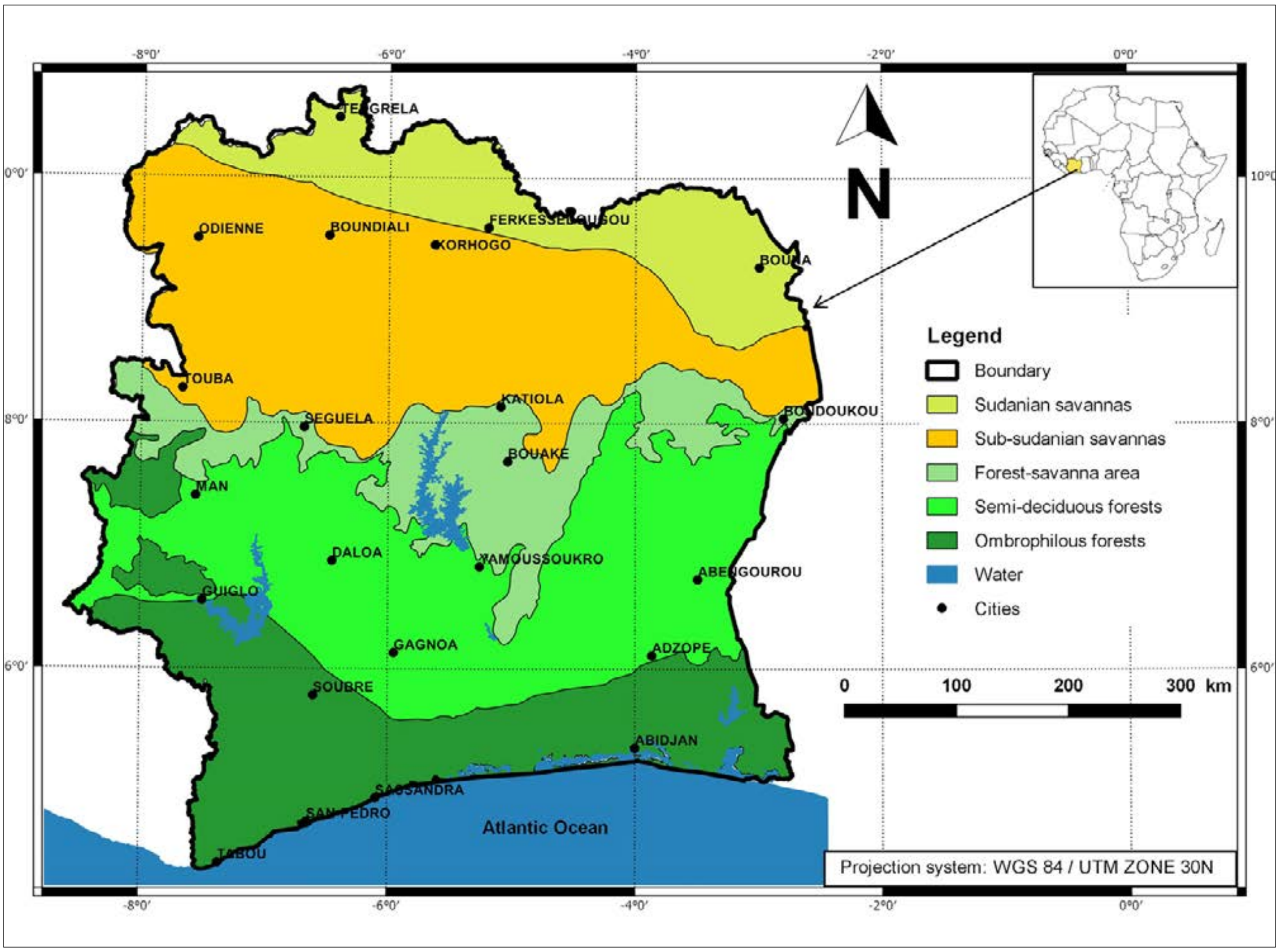


Figure 1: Study area defining the main vegetation types.

Table 1: Human population densities by region of Côte d'Ivoire.[15]

Region	Ecoregion	Vegetation type	Density (inhabitants/km²)
Abidian District	Guinean	Ombrophilous forest	2221.52
Agnéby-Tiassa	Guinean	Ombrophilous forest	66.83
Belier	Guinean	Forest-savanna	50.93
Cavally	Guinean	Ombrophilous forest	40.73
Gbôklé	Guinean	Ombrophilous forest	55.47
Gôh	Guinean	Semi-deciduous forest	119.57
Grands ponts	Guinean	Ombrophilous forest	54.43
Guemon	Guinean	Semi-deciduous forest	118.63
Haut-Sassandra	Guinean	Forest-savanna	80.57
Iffou	Guinean	Semi-deciduous forest	34.86
Indénié-Djuablin	Guinean	Semi-deciduous forest	80.99
Lôh-Djiboua	Guinean	Ombrophilous/semi-deciduous forest	82.95
Marahoué	Guinean	Forest-savanna	99.23
Mé	Guinean	Ombrophilous/semi-deciduous forest	62.49
Nawa	Guinean	Ombrophilous forest	114.55
N'Zi	Guinean	Semi-deciduous forest	66.55
Moronou	Guinean	Semi-deciduous forest	52.87
San-Pedro	Guinean	Ombrophilous forest	64.63
Sud-Comoé	Guinean	Ombrophilous forest	84.26
Tonpki	Guinean	Ombrophilous/semi-deciduous forest	80.80
Yamoussoukro District	Guinean	Semi-deciduous forest	101.59
Bafing	Guinean/Sudanian	Forest-savanna/Sudanian savanna	21.16
Béré	Guinean/Sudanian	Forest-savanna	29.32
Gbêkê	Guinean/Sudanian	Forest-savanna	110.64
Worodougou	Guinean/Sudanian	Forest-savanna	24.50
Gontougo	Guinean/Sudanian	Sub-sudanian savanna	39.78
Hambol	Guinean/Sudanian	Sub-sudanian savanna	22.49
Bagoue	Sudanian	Sub-/Sudanian savanna	35.22
Bounkani	Sudanian	Sudanian savanna	12.09
Folon	Sudanian	Sudanian savanna	13.32
Kabadougou	Sudanian	Sub-/Sudanian savanna	13.81
Poro	Sudanian	Sub-/Sudanian savanna	57.00
Tchologo	Sudanian	Sudanian savanna	26.40

Vegetation and structure

The dominant vegetation types, as delineated by Guillaumet and Adjanohoun[24], are shown in Figure 1. Ombrophilous forests in the south of the country consist of different types of evergreen forests. The canopies are the tallest, with some trees up to 40 m tall, and with total canopy cover of 100% in some places. The herbaceous stratum is diffuse. Semi-deciduous forests are humid forests with a structure comparable to ombrophilous forests, but with a sparser canopy cover. The herbaceous stratum varies from 1 m to 2 m tall. In the forest-savanna area, the landscape has a combination (a patchwork) of dry forest and humid savanna. Gallery forest and island forest are frequent. *Borassus aethiopum* savanna is the most common shrubland savanna. The coverage of the shrubby and woody strata can reach 50% and 35%, respectively. The herbaceous stratum is nearly continuous, with up to 90% coverage over some places, providing fuel favourable to fires. Sub-Sudanian and Sudanian areas are located in the north over the eighth parallel. Although the Sudanian area is the driest part of the country, these two areas have similar vegetation characteristics, mostly dry savannas. Some gallery and island forests are also present with a taller woody stratum. The important herbaceous stratum (dominated by Andropogoneae grasses) and the long dry season lead to frequent fires here.

MODIS fire data source

One of the sensors in the Earth Observation System, MODIS is a multi-temporal remote sensing device. The active fire data used in this study are derived from MODIS imagery and provided by the Fire Information for Resource Management System (FIRMS) of the US National Aeronautics and Space Administration (NASA) and the University of Maryland (USA). The MODIS instrument is equipped with Terra and Aqua satellites providing observations of Earth's surface four times per day: at 10:30 and 22:30 local time for Terra and at 13:30 and 01:30 for Aqua (times are for the Côte d'Ivoire). Thus, fire pixels are recorded four times per day, during the time of the satellite's overpass.

We used MODIS observations ranging over 10 years, from the start of the 2006/2007 fire season to the end of the 2015/2016 fire season. Thus, data were acquired according to fire seasons. Hereafter, each fire season is indicated using only the final year (as fire seasons span two calendar years). For example, the 2006/2007 season will be referred to as the 2007 fire year. The data cover all 31 regions and two autonomous districts of Côte d'Ivoire, i.e. the whole territory.

The original data were daily fire products with full spatial resolutions of 1 km (until 2010), 500 m (between 2010 and 2015) and 375 m (in 2016). Because these fire products were unmanageably large to be used in their entirety, they were plotted into fixed grids of 0.5°x0.5° in size in which fire counts were made.[25] Using the algorithm developed by Giglio et al.[16], MODIS data are processed using the Rapid Response System. The processed data are made available through the Joint Research Centre (JRC) web server (http://firetool.jrc.ec.europa.eu/) of NASA's FIRMS. Monthly maps of fire occurrence are available for download in shapefile (.shp) format and can be used in a GIS environment.

Data processing

Fire data acquired were imported in QGIS software in which fire pixels were overlaid with the shapefile of Ivorian administrative boundaries. Water bodies were masked out to avoid false alarms and ensure the accuracy of detected fire hotspots, as suggested by Giglio[25]. Once processed in QGIS, attribute tables of fire data were exported for descriptive statistical analysis. The spatio-temporal distribution of fire was investigated, based on the number of fire days, the number of active fires and fire density.

For all parameters (number of fire days, number of active fires and fire density), any value that was higher than the national average was designated by 'above-average' and a lower value by 'below-average'. Regions with above-average fires were considered as high-risk areas or fire-prone areas (relative to averages). Fire activity was tested for significance using Pearson's correlation coefficient test in R software.[26]

Active fires and number of fire days

An active fire is any fire detected by the satellite when it is still burning. Whenever a pixel was recorded with an active fire, that designation of a fire occurrence (or a fire pixel) was counted just one time during the relevant fire season. Active fires were recorded per region for each year (each fire season). Next, the spatial variability of the fire occurrences was mapped by administrative subdivisions of Côte d'Ivoire (i.e. 31 regions and two autonomous districts). Temporal variability was calculated by years and on a monthly basis. The map of active fires was produced using only fire pixels with a confidence level of 100% (high confidence), in order to better focus on fire hotspots (Figure 2).

A fire day was defined as a day on which at least one active fire (i.e. fire pixel) was detected by the MODIS satellite. Thus, the total number of fire days, defined as the sum of the fire days, was determined by year and by administrative subdivision.

Fire density and fire-prone regions

Fire density is defined as the number of fire pixels per 1000 ha detected over a given period of time.[20] It was analysed spatially and temporally, and is presented here as fires/km²/year. Based on the numerical size of fire density, fire-prone regions were identified. A 'fire-prone' region is defined as a region with above-average fires over the study period. For more accuracy on fire dynamics over main fire hotspots, the fire density of the peak region (region with highest fire density in fire hotspot) for a given fire season was assessed with regard to annual rainfall of the previous rainy season. Analyses included diachronic (i.e. the variation over time, through the entire study period) and synchronic (i.e. the variation in each main fire hotspot for the same period) evaluations.

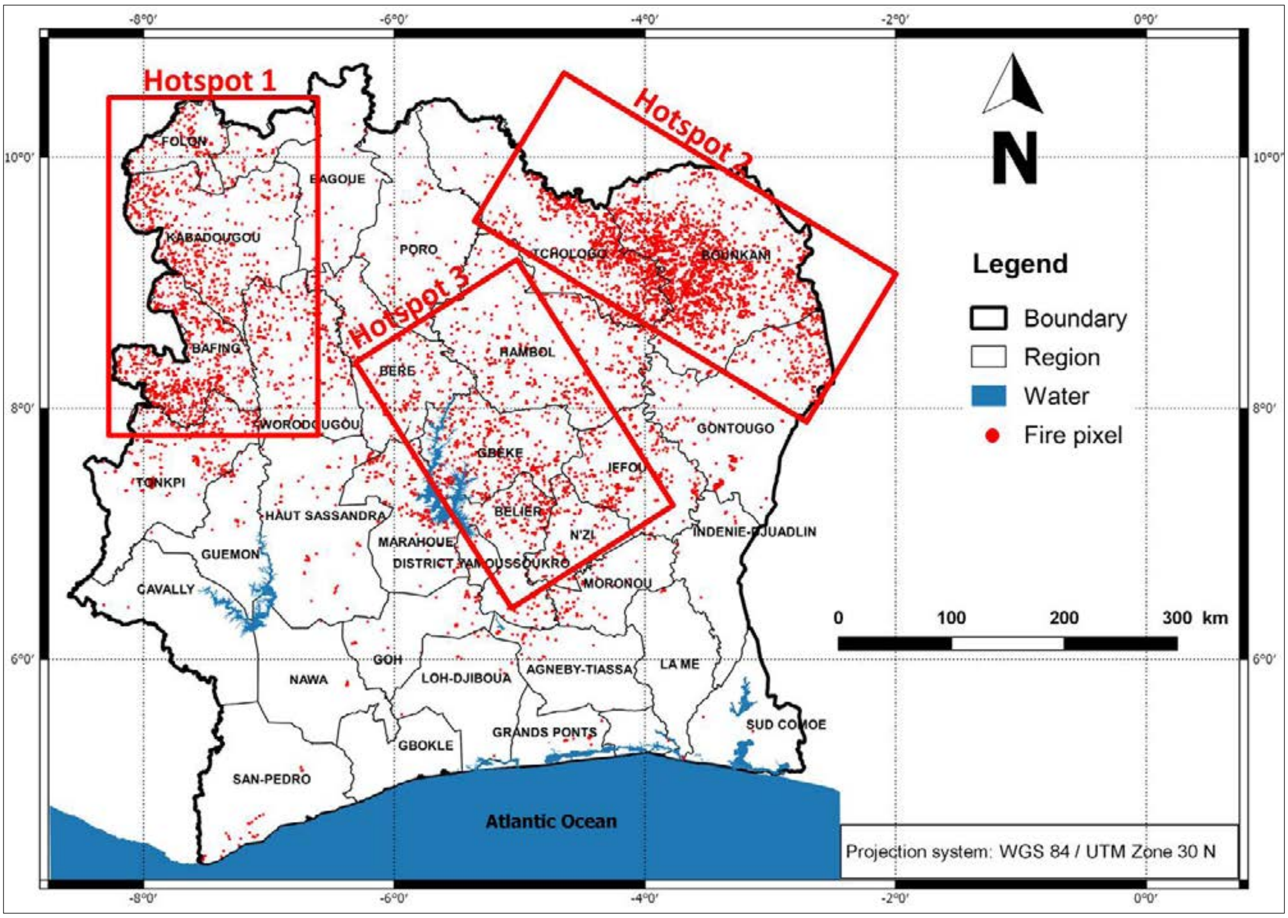


Figure 2: Map of the spatial distribution of active fires in Côte d'Ivoire showing main fire hotspots.

Potential relationships of fire variables with annual precipitation values were also explored. Precipitation data (from 2006 to 2015) was provided by the national meteorology company (Sodexam) and was used to investigate links between fire dynamics and rainfall in main fire hotspots.

Results

Fire spatial activity

Active fires spatial distribution

Active fires were more abundant in the North than the South, visible on the fire map made with 100% confidence fire pixels (Figure 2). The number of active fires ranked in decreasing order revealed regions with higher fire occurrence (Figure 3). Over the 10 years of the study, the country average number of MODIS active fires was 1279±95. The average fire number varied across administrative regions, from 78±11 (Gboklé region, South) to 6047±335 (Bounkani region, North-East). Also, regions such as Bounkani, Tchologo, Hambol, Gontougo (in the North-East), Kabadougou, Bafing, Worodougou, Folon, Tonkpi (in the North-West), Poro (in the North) and Gbêkê, Béré, Bélier (in the Centre) recorded above-average fires. For the Southern regions that had below-average fires, most recorded more than 100 active fires per year in average (Figure 3). This was the case of Gôh (281±54), San-Pedro (253±36), Lôh-Djiboua (190±32), Guémon (150±26), Sud-Comoé (139±14), Grands Ponts (136±15), Cavally (125±17), Indénié-Djuablin (120±15) and Mé (109±15).

Fire density variations

The national average fire density was 0.11±0.01 fires/km²/year. Fire densities varied greatly by region, however. When ranked in descending order using the 10-year data set (Figure 4), the Nawa region located in the South-West, had the lowest fire density (average 0.01±0.00 fires/km²/year) whereas the Bafing region in the North-West recorded the highest (0.4±0.02 fires/km²/year). Based on the comparison to the national average, among the 33 administrative subdivisions of the country, only 11 (located between 2°30'–8°30'W and 7°00'–10°30'N) were identified as prone to fire.

From these results, the 11 fire-prone regions were grouped into three main fire hotspots as high-risk contiguous areas. These three main fire hotspots were designated (from highest to lowest fire densities) as: North-Western, North-Eastern and Central hotspots (Figure 2). The North-Western (also designated as hotspot 1) is in Sub-Sudanian and Sudanian savannas, and includes the regions of Bafing, Kabadougou, Folon and Worodougou. The North-Eastern (hotspot 2) is in Sudanian savannas and includes the regions of Bounkani and Tchologo. The Central (hotspot 3) is located between Sub-Sudanian and forest-savanna areas and includes the regions of Hambol, Gbêkê, Bélier, Béré and Yamoussoukro district (Figure 2).

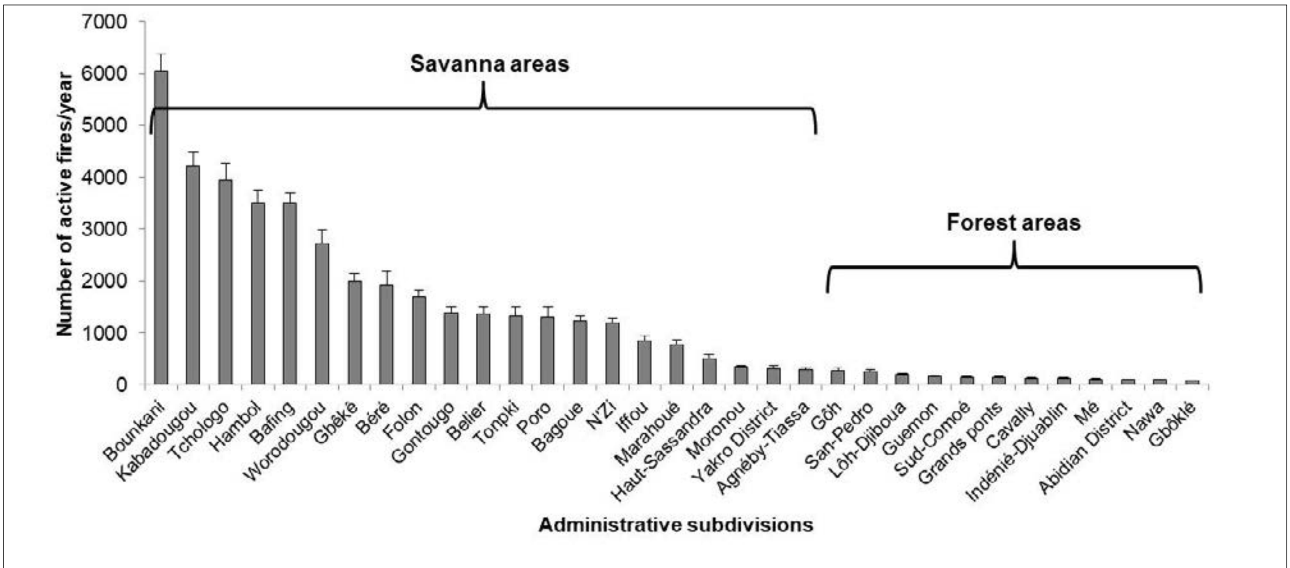


Figure 3: Average number of active fires per year in administrative subdivisions of Côte d'Ivoire over a 10-year period (2007 to 2016).

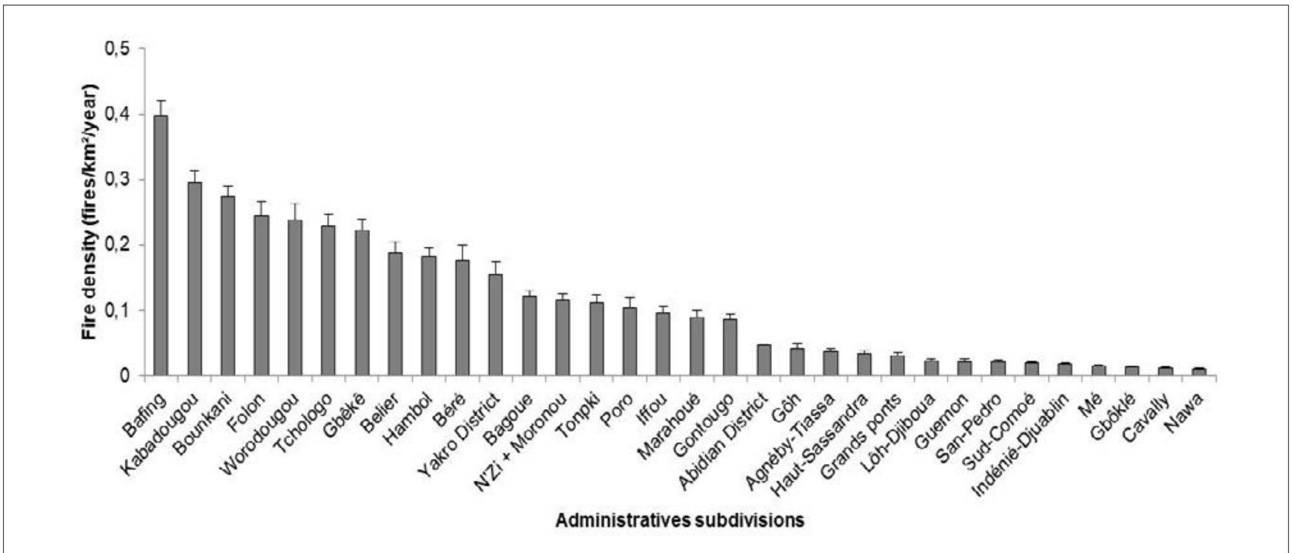


Figure 4: Average fire density per year in administrative subdivisions of Côte d'Ivoire over a 10-year period (2007 to 2016).

Fire temporal activity

Inter-annual variability of fires

At a national level, fires are present in Côte d'Ivoire about 205 days in a year. Over the 10 years of the present study, there was a slight decrease in the total number of fire days (Figure 5a; slope = about -2 days/year, t=-2.77, p=0.024, significant at 95% confidence level). Nevertheless, above-average numbers, i.e. higher numbers of fire days than the national average (205±3 days), were recorded in 2008, 2009, 2010 and 2012. The maximum of 217±5 and the minimum of 191±4 fire days were recorded in 2009 and 2016, respectively.

The number of active fires showed a similar trend to the number of fire days, with a steeper slope (Figure 5b; slope = about -2918 fires/year, t=-5.63, p=0.0004, significant at 95% confidence level).

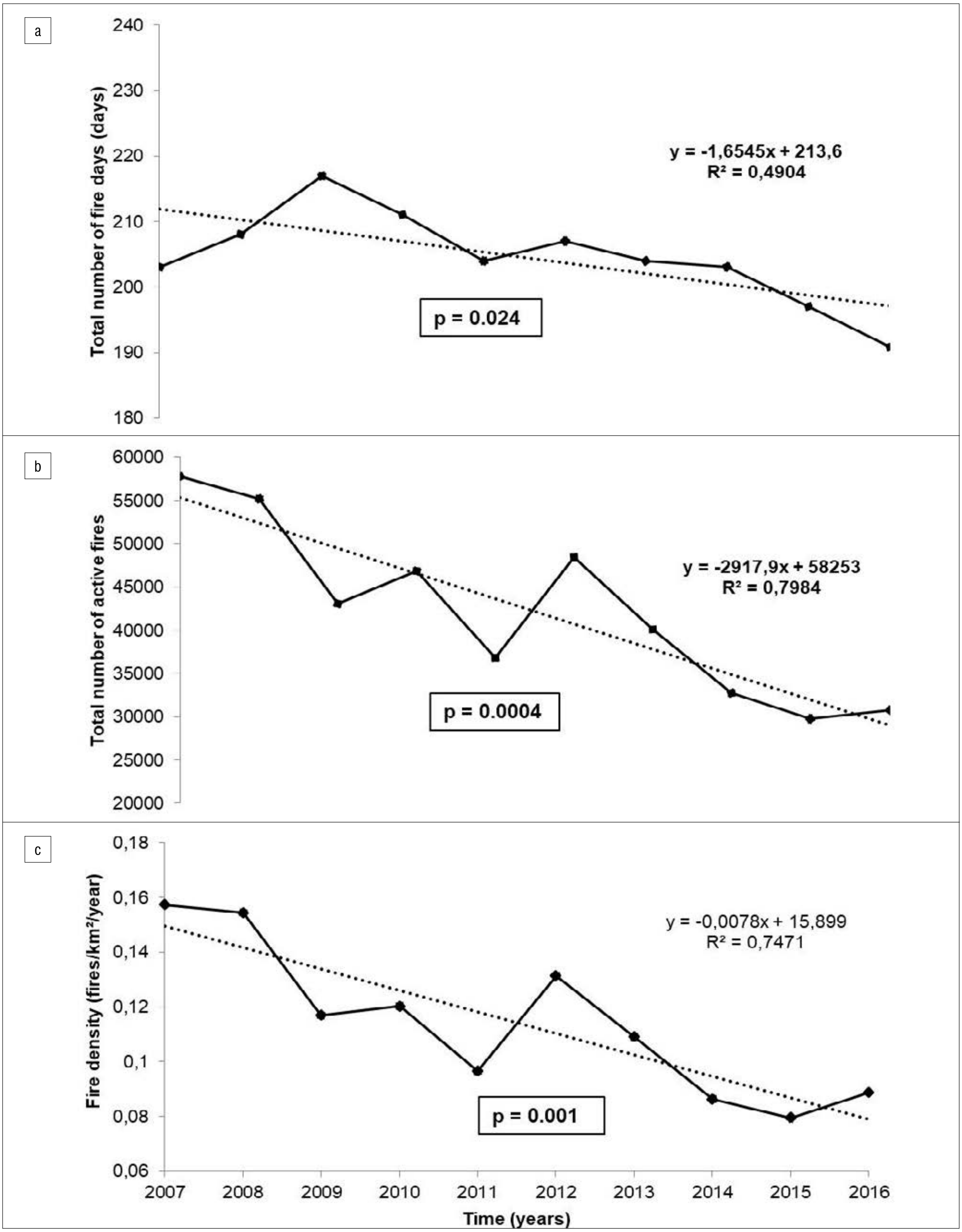


Figure 5: Yearly fire events in Côte d'Ivoire from 2007 to 2016: (a) total number of fire days; (b) total number of active fires; (c) fire density per area.

Over the 10 years, total active fires varied from 29 838±207 fires (2015) to 57 851±330 fires (2007), with 42 205±3226 fires as the annual average. A return time from 1 to 4 years was determined between years with above-average active fires (Figure 5b).

In general, the years from 2007 and 2010 had above-average numbers of active fires, whereas the years between 2013 and 2016 had below-average numbers. In 2011 (the middle of the study period), a below-average number was also recorded, although this number increased the following year (2012) to higher than the national average (Figure 5b). The highest fire peak and the lowest were observed during the 2007/2008 and 2015/2016 seasons, respectively (26 266 and 11 005 active fires, respectively).

For fire density, a similar trend was observed (Figure 5c; slope = about -0.01 fires/km²/year, t=-4.91, p=0.001, significant at 95% confidence level). There was also a return time from 1 to 4 years with above-average values. Fire density varied from 0.08±0.02 fires/km²/year (2015) to 0.16±0.02 fires/km²/year (2007), with a national average fire density of 0.11±0.01 fires/km²/year.

Fire seasonality

For the country as a whole, fires occur across a range of 7 months, from October to April, during the dry season (Figure 6). The same fire period was found in two of the hotspots that we identified: the Bafing (North-West) and Hambol (Centre) regions, whereas it was shortened from October to March in Bounkani (North-East) region (Figure 7).

Most fires occur between December and February, at both the national level and in each of the three fire hotspots. Fire occurrence during these three months was about 91% in Bafing region as at the national level, while it reached 95% and 96% in Bounkani and Hambol regions, respectively.

At national level, the peak month of fires was January (Figure 6), representing 39% of fire annual occurrence vs. about 36% in December and 16% in February. In Bafing region (North-West), fire peak was also in January with 41% of occurrence vs. 27% in December and 23% in February (Figure 7a). In contrast, Bounkani and Hambol regions (North-East and Central) experienced their fire peaks a bit earlier, in December (Figure 7b and 7c, respectively). Peak values were 62% in December vs. 26% in January and 7% in February for Bounkani, and 44% in December vs. 38% in January and 15% in February for Hambol.

Relation between fire density and rainfall in the main fire hotspots

Fire density peaks for the North-Western, North-Eastern and Central hotspots are shown in Figure 8: Bafing (Figure 8a), Bounkani (Figure 8b) and Hambol (Figure 8c). The average fire densities are 0.4±0.02, 0.28±0.02 and 0.18±0.01 fires/km²/year, respectively, for the North-Western, North-Eastern and Central fire hotspots. All three of these hotspots recorded above-average fire density, compared to the national yearly average of 0.11±0.01 fires/km²/year.

Further, the synchronic analysis (fire density and rainfall of the same year) within the main fire hotspots revealed that fire density was negatively correlated to annual rainfall overall, between 2007 and 2016 (Figure 8). That is, low rainfall resulted in high fire density. For each individual site, correlations between low rainfall and high fire density were significant for Bafing region only (Figure 8a; r=-0.96, p<0.0001). This significant negative correlation occurred in Bafing region where the 10-year average rainfall (1046 mm/year) was higher than that of Bounanki and Hambol regions (897 and 886 mm/year respectively).

Importantly, however, fire density was positively correlated with rainfall in the previous years, using diachronic analysis of rainfall from 2006 to 2015 with the fire density from 2007 to 2016 (Figure 9).

Discussion

Fire spatial distribution

The decreasing fire occurrence from the northern to southern regions of Côte d'Ivoire are most likely functions of gradients in rainfall and vegetation distributions from Sudanian (North) and Guinean (Centre) savannas to evergreen forests (South).[24] Indeed, the northernmost savanna areas are dominated by a well-developed herbaceous stratum which constitute the main fuel during fires. Such a relationship of precipitation, vegetation structure, herbaceous fuels and fire occurrence is similar to the demonstrated roles of vegetation structure and fuel chemistry in fire regimes in South Africa by Van Wilgen et al.[27]

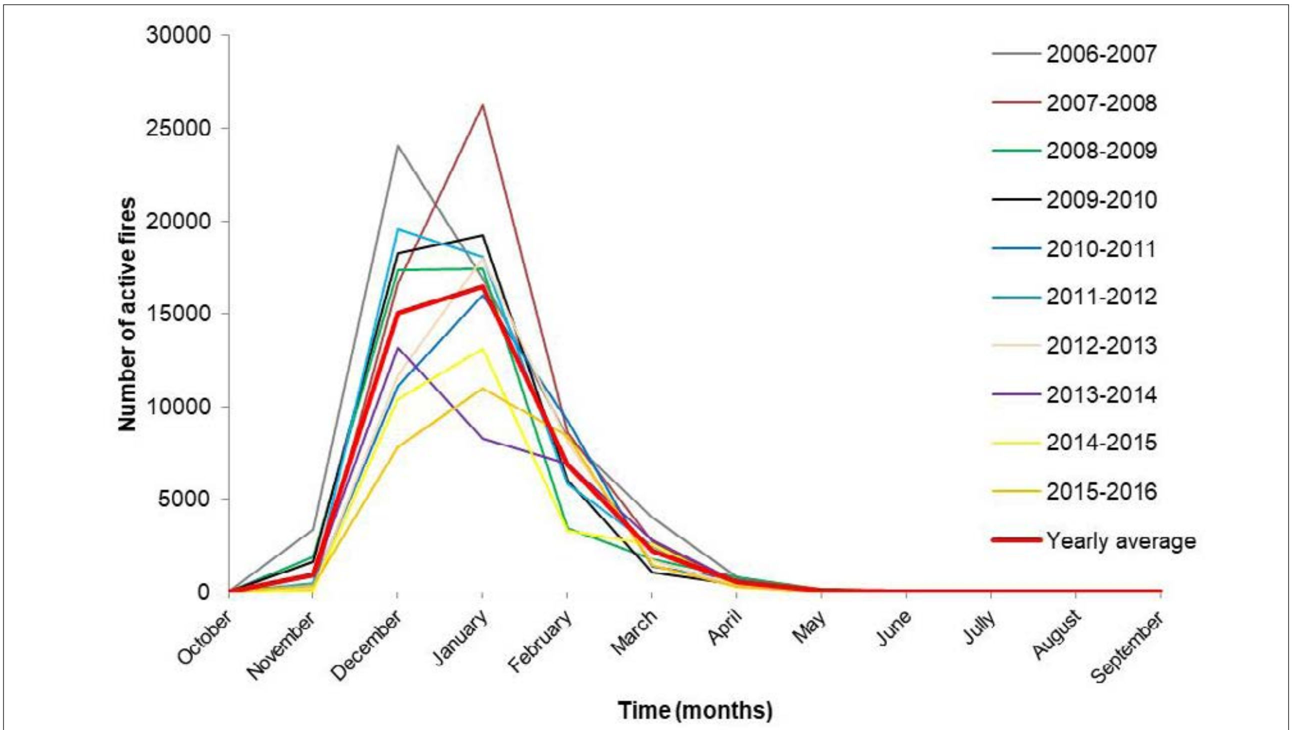


Figure 6: Temporal distribution of active fires in Côte d'Ivoire by month from 2007 to 2016.

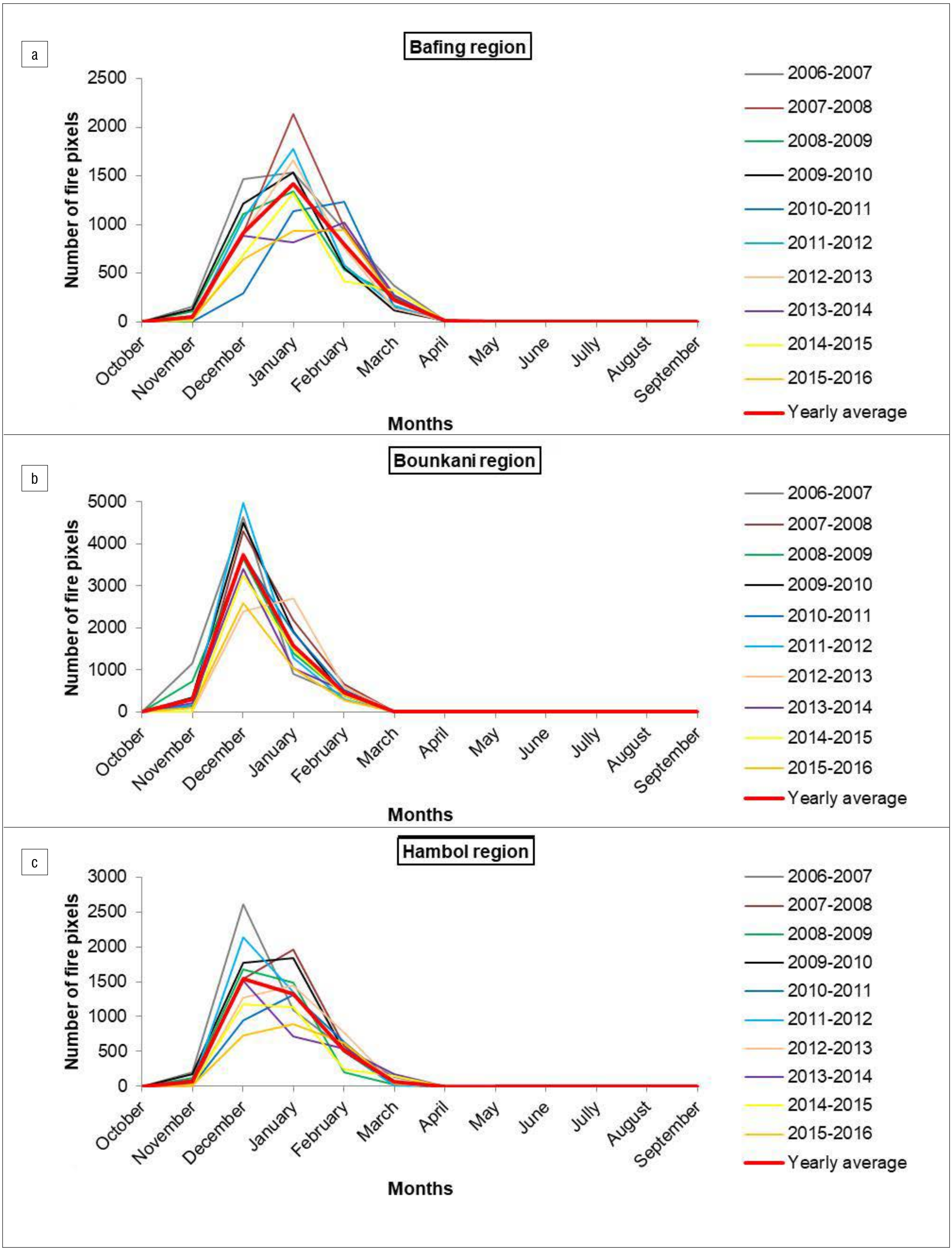


Figure 7: Temporal distribution of fire occurrence by month in the three main fire hotspots: (a) Bafing, main region of the North-Western hotspot; (b) Bounkani, main region of the North-Eastern hotspot; (c) Hambol, main region of the Central hotspot.

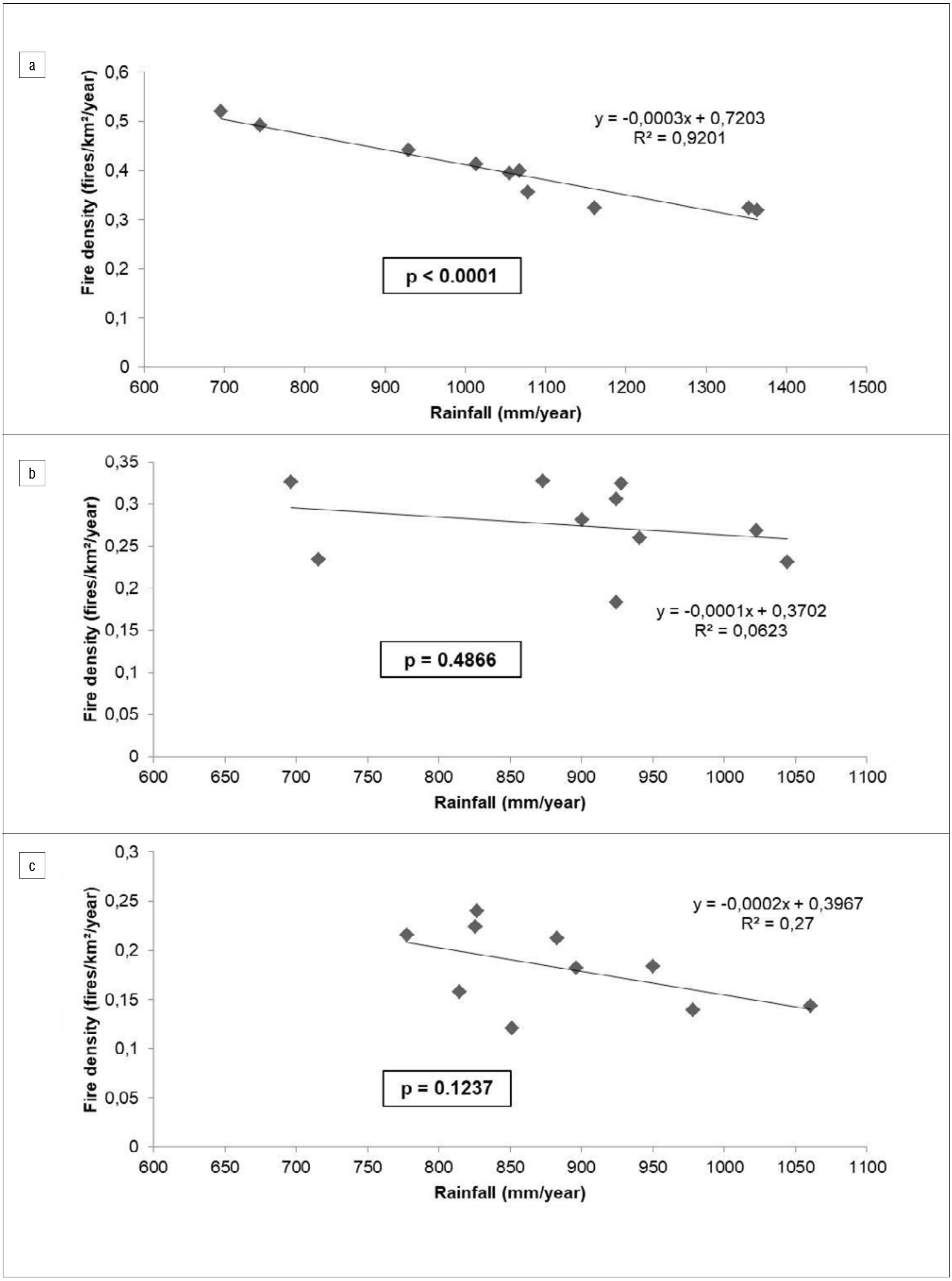


Figure 8: Synchronic variation of fire density relative to precipitation in the three main fire hotspots from 2007 to 2016. (a) Bafing, main region of the North-Western hotspot; (b) Bounkani, main region of the North-Eastern hotspot; (c) Hambol, main region of the Central hotspot.

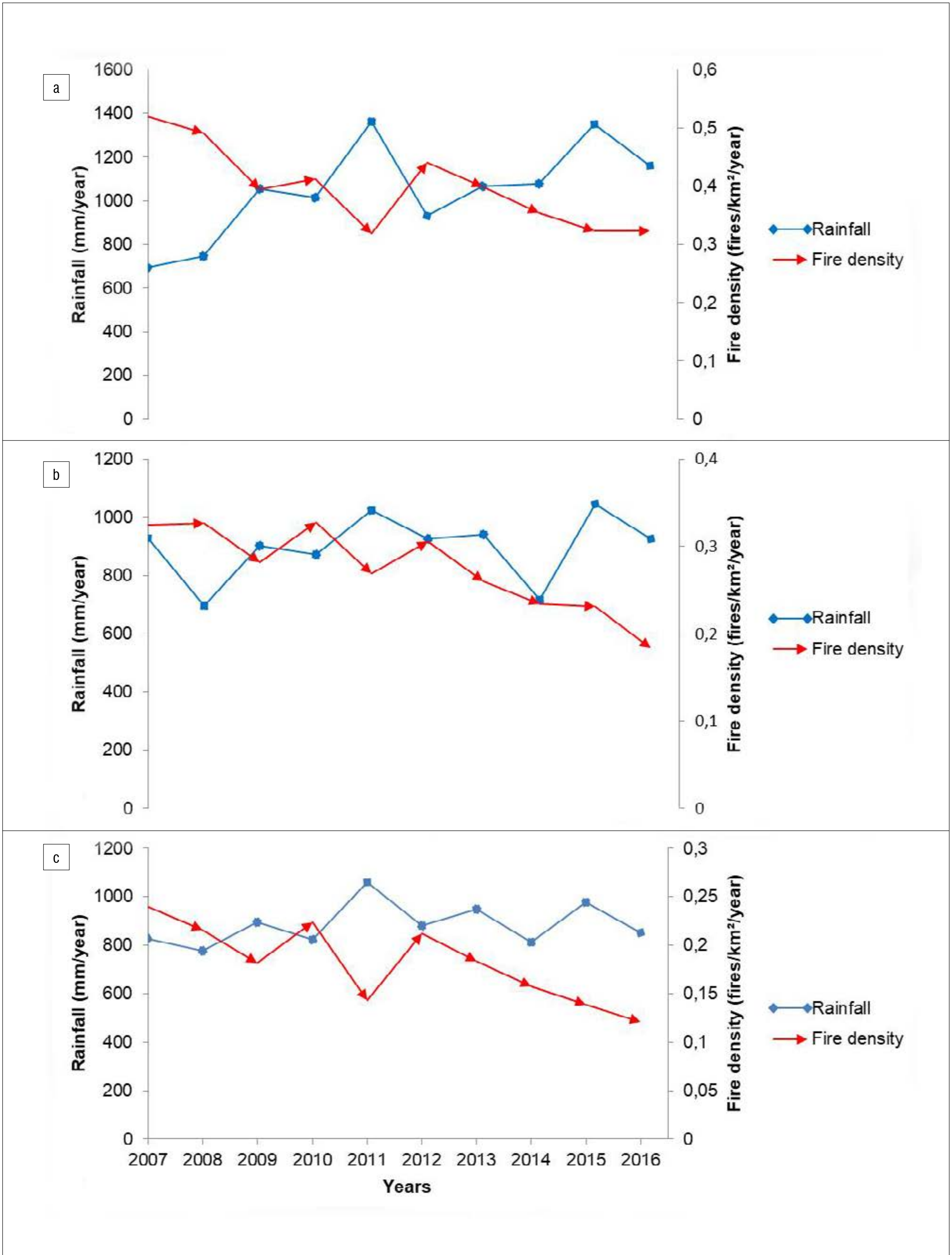


Figure 9: Diachronic variation of fire density (2007 to 2016) and precipitation in the previous wet season (2006 to 2015) in the three main fire hotspots: (a) Bafing, main region of the North-Western hotspot; (b) Bounkani, main region of the North-Eastern hotspot; (c) Hambol, main region of the Central hotspot.

The geographical locations of the three main fire hotspots we identified in Côte d'Ivoire are consistent with a regional study of fires in West Africa by N'Datchoh et al.[11], using burned area from the SPOT VEGETATION product between 2000 and 2007. In their map of vulnerability to fire over West Africa, they stressed that fires were mostly important over Sudanian and Guinean savannas in areas located in southeast Burkina Faso and the northern part of the Côte d'Ivoire–Ghana border (in which the Bounkani region lies). Our study used data over 10 years and fire pixels with 100% confidence of detection over the entire country, so we were able to explore local high-risk areas at greater depth.

The three main fire hotspots of Côte d'Ivoire are spatially distributed in a V-shape, similarly to Ivorian savanna's spatial distribution known as the V-Baoulé.[24] Thus, the heterogeneity of the fire spatial distribution is strongly related to the ecoregions of Côte d'Ivoire. Indeed, the North-Eastern fire hotspot is located in the Sudanian area where grasslands and shrublands are predominant, while the North-Western hotspot is extended from Sudanian to mountain areas composed of woodland savannas and mountain forests. The Central hotspot is located between sub-Sudanian and mesophilous areas in which humid savannas and semi-deciduous forests are mixed.[24]

No fire hotspot was found in the ombrophilous area dominated by evergreen forests. However, this does not mean that this area is sheltered from accidental fires. The most forest regions such as Gôh, San-Pedro, Lôh-Djiboua, Guémon, Sud-Comoé, Grands Ponts, Cavally, Indénié-Djuablin and Mé recorded more than 100 active fires per year on average, despite the fact that forest canopy, cloud cover or thick smoke can completely mask a fire so that it may not be detected by MODIS.[25] This fact presents a challenge to forest protection services which are charged with preserving the remaining forests of the country. They have to use other methods to facilitate large-scale monitoring to reinforce forest surveillance.

Temporal variations of fire activity

Fires occurred in Côte d'Ivoire over more than half the year (205 fire days, 56% of the year). The 7-month fire period for Côte d'Ivoire is much longer than the 93 fire days in Togo (West Africa) reported by Afelu[28] for the period between 1995 and 2014. We suggest the difference may be due to two factors: (1) fire management policies of Togo identify appropriate periods and dates of biomass burning according to ecological regions each year whereas such policies do not exist in Côte d'Ivoire; (2) the land area of Togo is one-sixth the size of Côte d'Ivoire and thus Côte d'Ivoire has a wider range of timing (onset and cessation) and length of fire seasons. Such differences and results here suggest that considerable efforts are needed to develop effective fire management policies in Côte d'Ivoire.

The fact that the maximum number of both active fires and fire densities were recorded during the 2006/2007 season could be related to large-scale climate factors (as well as local conditions). In the decade considered in our study, the year 2006 was classified as the warmest La Niña year.[13] Such an event produces increased air temperatures and higher rainfall which, in some areas, stimulates the growth of combustible biomass. High biomass production in one year has the potential to produce higher fire occurrences during the following year or soon thereafter.[29]

Indeed, in our study, inter-annual variations in both number of active fires and fire density showed a return time of 1 to 4 years with above-average values, meaning high fire activity in the country in these years. Such a return time may indicate the reconstitution rate of vegetation cover (i.e. combustible biomass) after intense fires.[28] Intense fires in a given year may decimate biomass, which might take 1 to 4 years to optimise its resilience and recover sufficient biomass for future intense fire occurrence. This finding is consistent with other studies that suggested that African savannas burn every 2 to 3 years.[11,30]

We do not know to what extent the high numbers of active fires recorded in our study could have been due in part to human-induced fires which increased during the socio-political crisis in Côte d'Ivoire from 2002 to 2011. During this time, particularly in the Central and Northern regions of the country, fire practices may have been intensified without fire management services to limit the local use of fires. Such a socio-political role in fire occurrences is also suggested by the decrease in fire activity and fire density after 2011 – a time which corresponds with the end of the socio-political crisis and a return to normal management conditions.

Several studies have predicted that global climate change will increase both the risk of extreme fires[12] and of fire frequency[13], due to changes in biomass availability and meteorological conditions during fires. In the long term, global climate must be taken into consideration when managing and preventing fires in Côte d'Ivoire. The study reported here will serve as a baseline in monitoring future changes in fire regimes.

Fire season and peaks

The 7-month fire season we describe for Côte d'Ivoire (October to April) is similar to the fire season described by Dwomoh and Wimberly[22] which occurred from November to May in West Africa. In both cases, there was some variation in seasonal patterns across ecoregions. Within this fire season, most fires (91%) occurred from December to February, which is consistent with the pattern reported by N'Datchoh et al.[11] who found that 95% of annual burned areas burned between November and February in West Africa. The single lengthy fire season is strongly associated with the weather patterns caused by the southward movement of the ITCZ (monsoon retreat) over the region.[11,23] In contrast, for South Africa, there are two main fire seasons per year, due to different climatic conditions.[13]

Within the long fire season described for the country as a whole, we identified three main fire hotspots when fire occurrence was the maximum between December and February. There were slight differences in timing and fire activity among the three hotspots, most likely due at least in part to different climatic conditions among the sites. The burned areas are mainly composed of grasslands and shrublands. Anthropogenic pressures exerted by human populations through slash-and-burn agricultural practices, livestock breeding, and hunting contribute to burning during December and January.[31] During these months (the heart of dry season), biomass dries more quickly, thus fostering rapid combustion.[31,32]

In the North-Eastern area, the dry season is very pronounced[23,24] and begins earlier than in the other hotspots; hence, the fire peak in December. The Central fire hotspot is located in a transition zone between the sub-Sudanian and forest-savanna areas, which may explain why fire activity shows characteristics with both the North-Western hotspot (length of fire season) and the North-Eastern hotspot (peak month). Although our study suggests that climate may regulate fire occurrence through precipitation and its effect on biomass production, weather parameters such as air humidity, air temperature and wind speed, which can influence fire propagation on a local level[11,31], were not addressed in this study.

Fire density in the main fire hotspots

Rainfall seemed to be the main driver of fire density (fires per area). Although fire is mainly an anthropogenic phenomenon in West Africa[11], climate regulates the quantity and state of biomass which is mainly the fuel for vegetation fires. The negative correlation between fire density and rainfall in synchronic analysis suggests that high rainfall leads to high moisture content of fuel, creating unfavourable conditions for rapid fire spread, and therefore less intense fires.[31] This negative correlation between fire density and rainfall of the same year could be caused by the influence of precipitation exceeding a certain threshold.[33,34] Indeed, Mbow et al.[34] found a precipitation threshold of about 800 mm/year, above which this influence is observed. In our study, the negative influence of rainfall on fire density was observed only in the Bafing region in forest-savanna area with about 1046 mm/year. In Bounkani and Hambol regions located in the Sudanian area where rainfall was relatively low (897 and 886 mm/year, respectively), this negative correlation was not observed. In our study, this rainfall threshold seemed to be about 900 mm/year.

It is certain that wet seasons control the biomass density and availability (fuel loads) which are limiting factors for fire propagation.[32] Further, we know that the relationship between fire during a dry season and precipitation is not linear.[11] Indeed, in our study, the diachronic analysis of rainfall with regard to fire density revealed that variations of fire density may depend on rainfall of the previous years. The best explanation for this finding was that the above-average values in fire statistics were observed from 1 to 4 years, which means that an increase in rainfall of the 1 to 4 preceding years resulted in an increase in fire activity. Similar findings were found for other African savannas, where rainfall in preceding years resulted in an increase in burnt area.[2,35]

In the three fire hotspots, fire density was higher than the national level, particularly in the North-Western and North-Eastern areas where fire densities were respectively more than triple and double the national average. Therefore, because high fire density can reflect the level of fire risk, fire impacts on regional and local scales can be severe and catastrophic, respectively, even if at global scale their effects are small.[10] This suggests that fire managers and decision-makers should focus on these high-risk areas (main fire hotspots) for better and optimal fire management resources provision during the dry season.

Conclusion

This study provides an insight into fire activity over the past decade in Côte d'Ivoire, identifying the high-risk areas and fire seasons. Three main fire hotspots were identified in savanna areas located in the North-West, the North-East and the Central parts of the country; thus, savanna areas burned more frequently than forest areas. The majority of fires in Côte d'Ivoire occur between December and February. Fire density seemed to be impacted by the amount of the previous year's rainfall with a return time of above-average fires (from 1 to 4 years) evident in fire statistics. A decreasing trend was observed in fire activity across the 10-year study period, most likely associated with temporary socio-economic factors. Indeed, the continuing extent and timing of the wildfire phenomenon require considerable efforts in fire management in Côte d'Ivoire. By identifying the main fire hotspots, their fire season and their fire peaks, this study can assist in the efficient allocation of financial, technical and human resources according to local conditions. The study may also assist decision-makers and institutions managing fires, such as the National Committee for Forest Protection and Fight against Wildfires, to prevent destructive bushfires. The study can serve as a base to assess the effects of both climatic parameters and vegetation types on fire dynamics in the fire hotspots of the Côte d'Ivoire.

Acknowledgements

We acknowledge Campus France, the Ministries of Europe and Foreign Affairs, and of the Higher Education, Research and Innovation (France) for funding through the 'Make Our Planet Great Again (MOPGA)' initiative (grant reference number: mopga-short-0000000184). We thank Professor Patricia Werner from Charles Darwin University (Research Institute for the Environment and Livelihoods, Darwin, Australia) and the anonymous reviewers for their valuable comments and suggestions that helped to improve the final manuscript.

Competing interests

We declare that there are no competing interests.

Authors' contributions

T.D.S.: Original formulation; design of methodology; data collection; data analysis; validation; data curation; writing – the initial draft; writing – revisions; funding acquisition. M.K.: Original formulation; design of methodology; validation; writing – revisions; student supervision; project leadership. A.B.N.: Original formulation; design of methodology; validation; writing – revisions; student supervision; project management. N.E.T.: Design of methodology; validation; writing – revisions.

References

1. Bond WJ, Woodward FI, Midgley GF. The global distribution of ecosystems in a world without fire. New Phytol. 2005;165(2):525–538. https://doi.org/10.1111/j.1469-8137.2004.01252.x
2. Archibald S, Nickless A, Govender N, Scholes RJ, Lehsten V. Climate and the inter-annual variability of fire in southern Africa: A meta-analysis using long-term field data and satellite-derived burnt area data. Glob Ecol Biogeogr. 2010;19(6):794–809. https://doi.org/10.1111/j.1466-8238.2010.00568.x
3. Bowman DMJS, Balch JK, Artaxo P, Bond WJ, Carlson JM, Cochrane MA, et al. Fire in earth system. Science. 2009;324:481–484. https://doi.org/10.1126/science.1163886
4. Tanpipat V, Honda K, Nuchaiya P. MODIS hotspot validation over Thailand. Remote Sens. 2009;1(4):1043–1054. https://doi.org/10.3390/rs1041043
5. Bailly Y. Maitriser les feux [Control fires]. Agripromo. 1992;4:5–6. French.
6. Langmann B, Duncan B, Textor C, Trentmann J, Van der Werf GR. Vegetation fire emissions and their impact on air pollution and climate. Atmos Environ. 2009;43(1):107–116. https://doi.org/10.1016/j.atmosenv.2008.09.047
7. N'Dri AB, Konan LN. Does the date of burning affect carbon and nutrient losses in a humid savanna of West Africa? Environ Nat Resour Res. 2018;8(3):102–116. https://doi.org/10.5539/enrr.v8n3p102
8. UNEP. Spreading like wildfire – Tropical forest fires in Latin America and the Caribbean : Prevention, assessment and early warning. Panama: United Nations Environment Programme; 2002; p. 96.
9. Zammit O. Détection de zones brûlées après un feu de forêt à partir d'une seule image satellitaire SPOT 5 par techniques SVM [Detection of burnt areas after a forest fire from a single SPOT 5 satellite image using SVM techniques]. Nice: Université Nice-Sophia Antipolis; 2008; p. 147. French. https://tel.archives-ouvertes.fr/tel-00345683
10. Cochrane MA. Fire in tropics. In: Cochrane MA, editor. Tropical fire ecology : Climate change, land use and ecosystems dynamics. Chichester: Springer; 2009. p. 1–18.
11. N'Datchoh ET, Konaré A, Diedhiou A, Diawara A, Quansah E, Assamoi P. Effects of climate variability on savannah fire regimes in West Africa. Earth Syst Dyn. 2015;6(1):161–174. https://doi.org/10.5194/esd-6-161-2015
12. IPCC. The physical science basis: Working Group I contribution to the fourth assessment report of the Intergovernmental Panel on Climate Change. Synthesis report. Geneva: IPCC; 2007. Available from: https://www.ipcc.ch/reports/
13. Strydom S, Savage MJ. A spatio-temporal analysis of fires in South Africa. S Afr J Sci. 2016;112(11/12), Art. #2015-0489. https://doi.org/10.17159/sajs.2016/20150489
14. Joseph MB, Rossi MW, Mietkiewicz NP, Mahood AL, Cattau ME, St. Denis LA, et al. Spatiotemporal prediction of wildfire extremes with Bayesian finite sample maxima. Ecol Appl. 2019;29(6), e01898. https://doi.org/10.1002/eap.1898
15. INS. Résultats du recensement général de la population et de l'habitat [Results of the general population and housing census]. Abidjan: Institut National de la Statistique; 2014. French. Available from: http://www.ins.ci/n/
16. Giglio L, Descloitres J, Justice CO, Kaufman YJ. An enhanced contextual fire detection algorithm for MODIS. Remote Sens Env. 2003;87(2–3):273–282. https://doi.org/10.1016/S0034-4257(03)00184-6
17. Justice CO, Giglio L, Korontzi S, Owens J, Morisette JT, Roy D, et al. The MODIS fire products. Remote Sens Env. 2002;83(1–2):244–262. https://doi.org/10.1016/S0034-4257(02)00076-7
18. Laris P. Integrating land change science and savanna fire models in West Africa. Land. 2013;2:609–636. https://doi.org/10.3390/land2040609
19. Le Page Y, Oom D, Silva J Jonsson P, Pereira J. Seasonality of vegetation fires as modified by human action: Observing the deviation from eco-climatic fire regimes. Glob Ecol Biogeogr. 2010;19:575–588. https://doi.org/10.1111/j.1466-8238.2010.00525.x
20. Grégoire JM, Simonetti D. Interannual changes of fire activity in the protected areas of the SUN Network and other parks and reserves of the West and Central Africa region derived from MODIS observations. Remote Sens. 2010;2:446–463. https://doi.org/10.3390/rs2020446
21. Yao N. Mapping bushfire distribution and burn severity in West Africa using remote sensing observations [PhD thesis]. Würzburg: University of Würzburg; 2010; p. 137.

22. Dwomoh F, Wimberly M. Fire regimes and their drivers in the upper guinean region of West Africa. Remote Sens. 2017;9(11):1117. https://doi.org/10.3390/rs9111117

23. Swap RJ, Annegarn HJ, Suttles JT, Haywood J, Helmlinger MC, Hely C, et al. The Southern African Regional Science Initiative (SAFARI 2000): Overview of the dry season field campaign. S Afr J Sci. 2002;98(3–4):125–130.

24. Guillaumet JL, Adjanohoun E. La vegetation [The vegetation]. In: Avenard JM, Eldin M, Girard G, Sircoulon J, Touchebeuf P, Guillaumet JL, et al., editors. Le milieu naturel de la Côte d'Ivoire [The natural environment of Côte d'Ivoire]. Mémoires ORTOM, Paris; 1971. p. 160–262. French.

25. Giglio L. MODIS Collection 6 active fire product user's guide revision A. College Park, MD: Department of Geographical Sciences, University of Maryland; 2015. p. 64.

26. R Development Core Team. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2014.

27. Van Wilgen BW, Higgins KB, Bellstedt DU. The role of vegetation structure and fuel chemistry in excluding fire from forest patches in the fire-prone fynbos shrublands of South Africa. J Ecol. 1990;78(1):210–222. https://doi.org/10.2307/2261046

28. Afelu B. Contribution à la gestion des feux de végétation au Togo [Contribution to the management of vegetation fires in Togo] [PhD thesis]. Lomé: Université de Lomé; 2016. p.177.

29. Koffi KF, N'Dri AB, Lata JC, Konaté S, Srikanthasamy T, Konaré S, et al. Effect of fire regimes on the demographic parameters of the perennial tussock grasses of a humid savanna. J Veg Sci. 2019;30(5):950–962. https://doi.org/10.1111/jvs.12788

30. Andreae MO. Biomass Burning: Its history, use and distribution and its impact on environmental quality and global climate. In: Levine JS, editor. Global biomass burning: Atmospheric, climatic and biospheric implications. Cambridge, MA: MIT Press; 1991. p. 3–21.

31. N'Dri AB, Soro TD, Gignoux J, Dosso K, Koné M, Koné NA, et al. Season affects fire behavior in annually burned humid savanna of West Africa. Fire Ecol. 2018;14:5. https://doi.org/10.1186/s42408-018-0005-9

32. Van Der Werf GR, Randerson JT, Giglio L, Gobron N, Dolman AJ. Climate controls on the variability of fires in the tropics and subtropics. Glob Biogeochem Cy. 2008;22:GB3028. https://doi.org/10.1029/2007GB003122

33. Archibald S, Lehmann CE, Gómez-Dans JL, Bradstock RA. Defining pyromes and global syndromes of fire regimes. Proc Natl Acad Sci USA. 2013;110(16):6442–6447. www.pnas.org/cgi/doi/10.1073/pnas.1211466110

34. Mbow C, Nielson TT, Rasmussen K. Savanna fires in East-Central Senegal: Distribution patterns, resource management and perceptions. Hum Ecol. 2000;28:561–583.

35. Archibald S, Roy DP, Van Wilgen BW, Scholes RJ. What limits fire? An examination of drivers of burnt area in Southern Africa. Glob Chang Biol. 2009;15:613–630. https://doi.org/10.1111/j.1365-2486.2008.01754.x

Check for updates

AUTHORS:
Andrew L. Skowno[1,2]
Debbie Jewitt[3,4]
Jasper A. Slingsby[5,6]

AFFILIATIONS:
[1]Kirstenbosch Research Centre, South African National Biodiversity Institute (SANBI), Cape Town, South Africa
[2]Plant Conservation Unit, Department of Biological Sciences, University of Cape Town, Cape Town, South Africa
[3]Biodiversity Research and Assessment, Ezemvelo KZN Wildlife, Pietermaritzburg, South Africa
[4]School of Animal, Plant and Environmental Sciences, University of the Witwatersrand, Johannesburg, South Africa
[5]Fynbos Node, South African Environmental Observation Network (SAEON), Centre for Biodiversity Conservation, Cape Town, South Africa
[6]Centre for Statistics in Ecology, Environment and Conservation, Department of Biological Sciences, University of Cape Town, Cape Town, South Africa

CORRESPONDENCE TO:
Andrew Skowno

EMAIL:
askowno@gmail.com

DATES:
Received: 17 Apr. 2020
Revised: 04 Aug. 2020
Accepted: 17 Aug. 2020
Published: 29 Jan. 2021

HOW TO CITE:
Skowno AL, Jewitt D, Slingsby JA. Rates and patterns of habitat loss across South Africa's vegetation biomes. S Afr J Sci. 2021;117(1/2), Art. #8182. https://doi.org/10.17159/sajs.2021/8182

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:
land-cover change, monitoring, biodiversity, ecosystem, indicator

FUNDING:
South African National Research Foundation (grant no. 118593)

Rates and patterns of habitat loss across South Africa's vegetation biomes

The loss of natural habitat resulting from human activities is the principal driver of biodiversity loss in terrestrial ecosystems globally. Metrics of habitat loss are monitored at national and global scales using various remote sensing based land-cover change products. The metrics go on to inform reporting processes, biodiversity assessments, land-use decision-making and strategic planning in the environmental and conservation sector. We present key metrics of habitat loss across South Africa at national and biome levels for the first time. We discuss the spatial patterns and trends, and the implications and limitations of the metrics. Approximately 22% of the natural habitat of South Africa has been lost since the arrival of European settlers. The extent and the rate of habitat loss are not uniform across South Africa. The relatively mesic Grassland, Fynbos and Indian Ocean Coastal Belt biomes have lost the most habitat, while the arid Nama-Karoo, Succulent Karoo and Desert have lost the least. Rates of loss increased across all biomes in recent years (2014–2018), indicating that the historical drivers of change (i.e. expansion of croplands, human settlements, plantation forestry and mining) are intensifying overall. We should caution that the losses we report are conservative, because the land-cover change products do not capture degradation within natural ecosystems. Preventing widespread biodiversity losses and securing the benefits we derive from biodiversity requires slowing and preventing further habitat degradation and loss by using existing land-use planning and regulatory tools to their full potential.

Significance:

- The loss of natural habitat resulting from human activities is the principal driver of biodiversity loss in terrestrial ecosystems in South Africa.
- Monitoring trends and patterns of habitat loss at a national scale provides a basis for informed environmental decision-making and planning, thus equipping civil society and government to address habitat loss and protect biodiversity while also meeting key development and socio-economic needs.

Introduction

The loss of natural habitat caused by human activities such as crop farming and infrastructure development is the principal driver of biodiversity loss in terrestrial ecosystems, globally and in South Africa.[1,2] These changes are increasingly reliably detected through the use of satellite remote sensing platforms[3], and there is a growing focus on developing tools to automatically detect change in near-real time[4,5]. The land-cover change products that result are now widely used as the basis for species and ecosystem risk assessments[3,6], land-use decision-making and strategic planning in the environmental and conservation sector[7,8]. The utility of these products stretches well beyond the environmental sector and forms a key source of intelligence for planning and management across multiple sectors including infrastructure development, agricultural planning, defence, health, mining and energy.[9]

In this study, we used recently released, high-resolution, national land-cover data from three time points (1990, 2014 and 2018)[10-12], and a baseline of 1750 before widespread anthropogenic land-cover change[13,14], to estimate rates of habitat loss across the whole of South Africa's land mass. Although a comprehensive land-cover change study has been undertaken at a provincial scale for KwaZulu-Natal[15], this is the first national-scale analysis for South Africa. We define habitat loss as the persistent loss of natural or near natural vegetation cover through anthropogenic activities and focus on the extent and rates of loss of natural terrestrial habitat as described in the *Vegetation of South Africa, Lesotho and Swaziland*[13,16]; including 458 distinct vegetation types, grouped into nine biomes plus azonal types. The Biodiversity Intactness Index[17] and South Africa's national biodiversity assessments[2] used vegetation types as the unit of assessment and clearly illustrated the wide application to biodiversity conservation efforts of land-cover change analysis.

Habitat loss is a key indicator in the Global Biodiversity Framework 2050, proposed by the Convention on Biological Diversity as the replacement of the Aichi Targets. To aid in reporting against the new Global Biodiversity Framework and the Sustainable Development Goals, we applied the recently described Ecosystem Area Index (EAI)[18] to our data as an overall indicator of the state of South Africa's terrestrial ecosystems. We then disaggregated the EAI to the biome level and explored the spatial patterns of habitat loss and the rates of change in different regions of the country. The implications of the observed changes for the country's unique and globally significant biodiversity are discussed.

Methods

Land-cover data sets for 1990, 2014 and 2018[10-12] were modified to improve the representation of abandoned croplands and artificial water bodies and combined onto a common reference grid to allow for pixel-level comparisons (see supplementary material). The data sets were reclassified to a simplified scheme with seven land-cover classes (Supplementary table 1), including one class for natural/near natural vegetation and six anthropogenic classes: (1) croplands, including all field and horticultural crops, irrigated and dryland; (2) built-up areas, including infrastructure and human settlements; (3) plantation forestry areas; (4) mining areas; (5) artificial water bodies and (6) secondary natural areas (previously ploughed, mined or developed areas, which have recovered some semblance of natural vegetation cover since abandonment).

The extent and rates of loss of natural vegetation cover were calculated for each of the 458 vegetation types delineated and described in the *Vegetation of South Africa, Lesotho and Swaziland*.[13,16] The total extent of habitat loss caused by the expansion of each of the six anthropogenic land-cover classes since European settlement was evident from the 2018 data set, while recent change was calculated by cross tabulating the 1990 and 2018 data sets. The remaining natural extent *(RnE)* of each vegetation type in 1990, 2014 and 2018 was calculated and expressed as a proportion of the original historical extent of the type prior to major land-cover changes in South Africa (i.e. circa 1750). The rate of recent habitat loss (*RoL*), between 1990 and 2018, was calculated and expressed as a proportion of the 1990 extent, divided by the difference in years between the time points. We then calculated the Ecosystem Area Index[18] for South Africa by calculating the geometric mean of the *RnE* values for all 458 vegetation types in the country, and did the same for each biome, based on the equation $EAI = \sqrt[n]{a_1 \cdot a_2 \cdots a_n}$, where n is the number of vegetation (ecosystem) types (458) and a is the proportion of natural habitat remaining for each ecosystem type at each time point.

Results

Extent of loss

In 1990, natural vegetation covered just over 80% of South Africa (1 019 005 km^2; Table 1). Between 1990 and 2018, a further 33 849 km^2 was cleared for various human activities, leaving just over 78% (951 831 km^2) of the original natural vegetation – a 3% decline over the 28-year period. Biome-level statistics revealed that the loss was not uniform, with the relatively mesic Indian Ocean Coastal Belt, Grassland and Fynbos biomes losing more of their original extent than others (Table 1, Supplementary tables 2 and 3). The establishment of crop fields, human settlements and plantation forestry are the dominant drivers of change in these regions (Figure 1, Supplementary tables 2 and 3).

Table 1: Remaining natural extent (RnE) per biome at each time point in the land-cover change time series. The table includes the rate of habitat loss (RoL) between 1990 and 2018, and between 2014 and 2018.

Biome	†1750 (km²)	1990 (km²)	%remaining 1990	2014 (km²)	%remaining 2014	2018 (km²)	%remaining 2018	Rate of loss (1990–2018)	Rate of loss (2014–2018)
Albany Thicket	35 307	32 486	92%	32 159	91%	31 988	91%	0.05	0.13
Desert	6262	6181	99%	6168	98%	6142	98%	0.02	0.10
Forests	4711	3987	85%	3896	83%	3882	82%	0.09	0.09
Fynbos	81 645	57 993	71%	55 958	69%	54 918	67%	0.19	0.46
Grassland	330 848	209 225	63%	198 044	60%	193 631	59%	0.27	0.56
Indian Ocean Coastal Belt	11 692	4882	42%	4195	36%	3945	34%	0.69	1.49
Nama-Karoo	249 354	245 220	98%	244 526	98%	244 192	98%	0.01	0.03
Savanna	394 102	327 806	83%	319 048	81%	316 490	80%	0.12	0.20
Succulent Karoo	78 208	74 910	96%	74 610	95%	74 386	95%	0.02	0.07
Azonal Vegetation	26 428	22 055	83%	21 564	82%	21 339	81%	0.12	0.26
Total	**1 218 557**	**984 743**	**80.8%**	**960 169**	**78.8%**	**950 913**	**78.0%**	**0.12**	**0.24**

†1750 is considered the historical baseline date, before widespread habitat loss began.

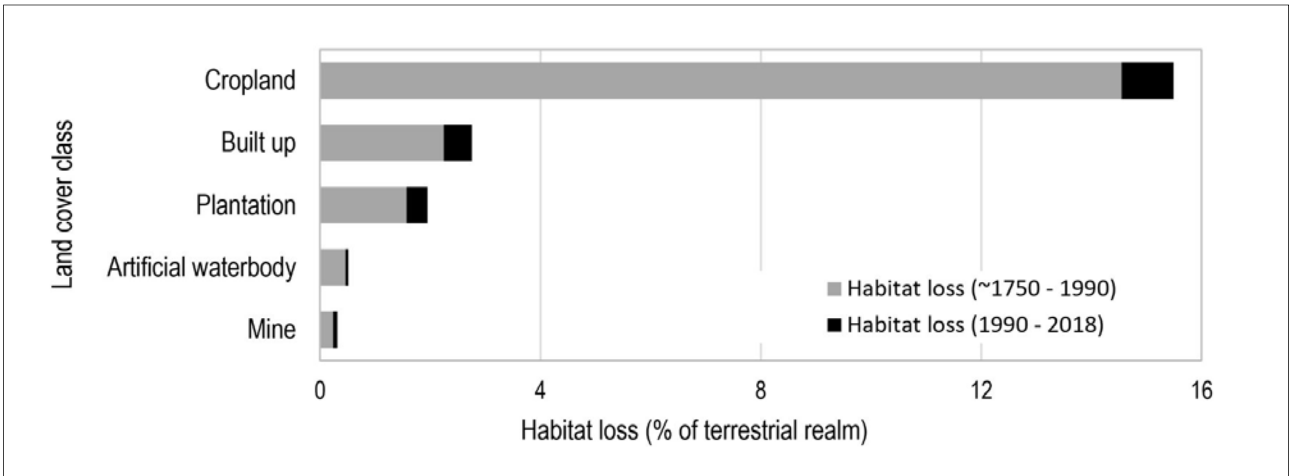


Figure 1: Habitat loss per major land-cover class (shown as a percentage of the terrestrial realm). Historical loss (between ~1750 and 1990) is shown in grey and recent loss (1990–2018) is shown in black. The clearing of natural habitat for the establishment of field and horticultural crops is the single largest historical driver of biodiversity loss in South Africa.

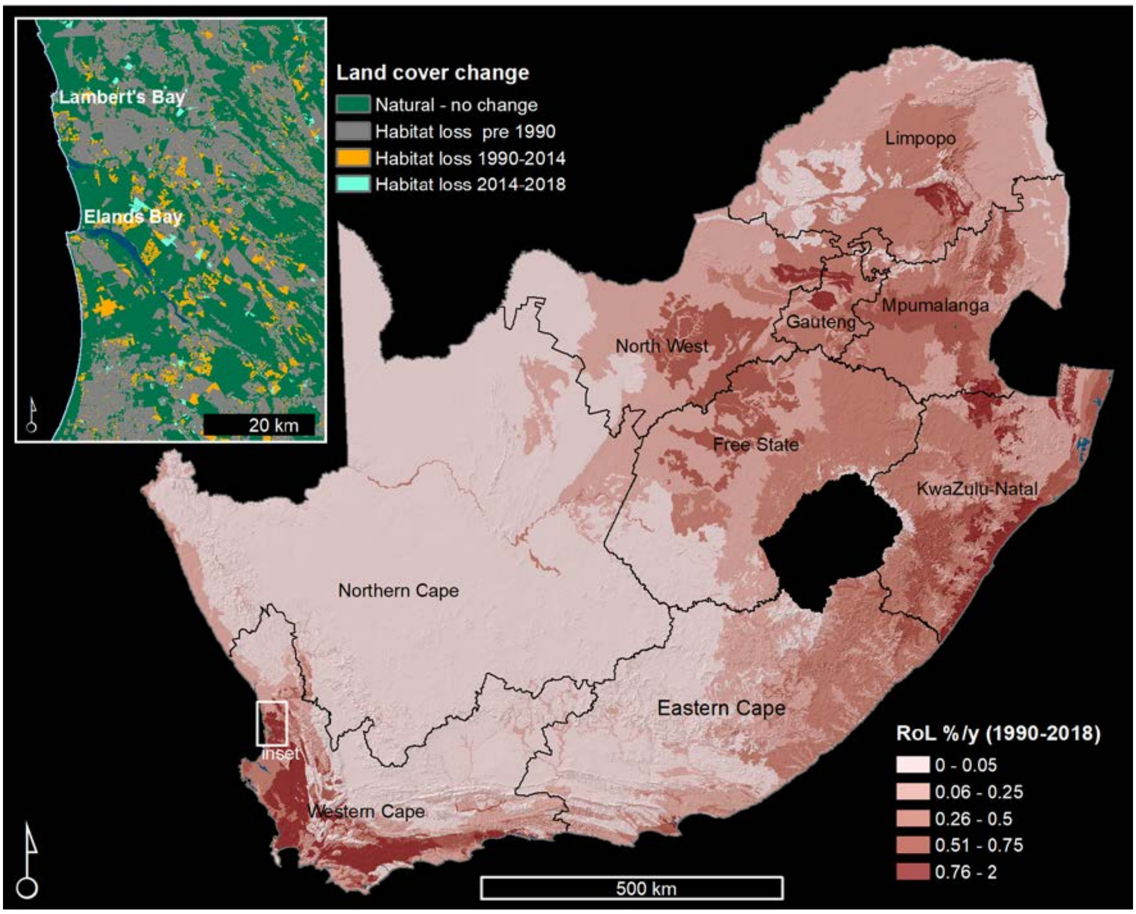


Figure 2: Rate of habitat loss indicator (RoL) for the period 1990–2018, calculated per terrestrial ecosystem type (vegetation types as defined by Mucina and Rutherford[13]). The Cape lowlands, Mpumalanga Highveld grasslands and KwaZulu-Natal coast and adjacent interior had the highest rates of habitat loss between 1990 and 2018, with expanding croplands and human settlements being the key drivers. The inset map shows the land-cover change for a coastal portion of the Western Cape (known as the 'Sandveld') where cropland expansion has occurred during each time period.

Rates of loss

Overall, South Africa lost 0.12% of its natural vegetation per year between 1990 and 2018, but this rate has been much faster in recent times – 0.24% per year between 2014 and 2018 (Table 1). As for the extent of loss, there are major differences in the rates of loss between biomes and between vegetation types (Figure 2). Rates of loss are highest in the southern and eastern coastal regions, in the interior Highveld grasslands and in the savannas of Limpopo and Mpumalanga. The arid interior had the lowest rates of loss. The key drivers of change are clearing for new croplands in the grasslands and expanding rural/peri-urban human settlements in the savannas in the northern parts of the country (Table 2, Supplementary tables 2 and 3). Loss of vegetation cover in coastal regions was linked to expansion of cultivation and human settlements.

Ecosystem Area Index

The Ecosystem Areas Index (EAI) for South Africa showed a 7% decline between 1990 and 2018, from 0.752 to 0.702 (Figure 3). In the Indian Ocean Coastal Belt biome there has been an 18% decline in the EAI between 1990 and 2018, despite already having less than 50% of its original extent by 1990. The arid Nama Karoo and Succulent Karoo biomes show less decline than more mesic biomes. The EAI for the Desert biome declined by 5% due to the high rate of habitat loss (due to coastal mining) in two small ecosystem types (Namib Lichen Fields and Alexander Bay Coastal Duneveld), despite the overall extent of natural habitat in the biome remaining relatively high (Table 1 and Supplementary table 3).

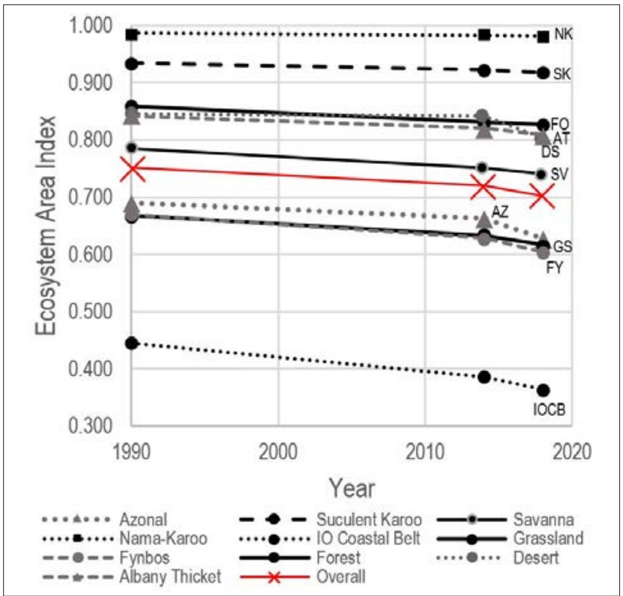


Figure 3: Ecosystem Area Index (EAI) for terrestrial ecosystems of South Africa: 1990, 2014 and 2018. There has been an overall decline of 7% across the whole of South Africa (1990–2018). For Fynbos, Grassland, Azonal vegetation and the Indian Ocean Coastal Belt, relative decline in EAI is most severe over the period from 2014 to 2018.

Discussion

The habitat extent and rate of loss metrics indicate that recent land-cover changes are impacting the same terrestrial ecosystems which suffered historical impacts, likely because the major drivers remain largely the same (i.e. croplands, human settlements, plantation forestry and mining). The increased rate of habitat loss in recent years (2014–2018) indicates that these drivers are intensifying. Variation in drivers between biomes has led to stark differences in habitat loss, such as between the relatively mesic Grassland, Savanna, Indian Ocean Coastal Belt and Fynbos biomes, and the arid Nama-Karoo, Succulent Karoo and Desert biomes. The Indian Ocean Coastal Belt, a narrow biome (~20 km wide) stretching from the Eastern Cape through KwaZulu-Natal into Mozambique, stands out in terms of high overall habitat loss and rate of loss. In this region in particular, it is crucial that the bioregional plans and biodiversity sector plans (which now cover the whole of South Africa's mainland) are fully considered in land-use and development planning.[19-21] These plans clearly identify a network of Critical Biodiversity Areas and Ecological Support Areas in which further habitat loss needs to be prevented, in order to secure crucial ecosystem services and allow biodiversity to persist. The focus of impacts on the relatively mesic regions may change in coming years, with emerging pressures such as renewable energy facilities (solar photovoltaic and wind in particular) and intensive wildlife breeding operations threatening arid biomes.[2,7] Forests remain relatively unchanged, likely due to a long history of legislated and cultural protections.[22,23]

Compared to direct habitat loss, the impacts of overgrazing, disrupted fire regimes and invasive species on biodiversity and ecosystem function are far more difficult to quantify.[1,24] A large portion of South Africa is subject to these pressures, but these changes were not detected or included in the national land-cover data set used in our analysis, and information on the overall condition of ecosystems that incorporates these elements is still lacking.[2] As such, the results presented here should be seen as a conservative estimate of the threats to South Africa's biodiversity. In particular, overgrazing is likely to alter the current estimates of >95% natural extent remaining for the arid biomes (Desert, Nama and Succulent Karoo), as much of this area is used for rangelands, while altered fire regimes and invasive species are likely to impact the more productive and flammable biomes (Fynbos, Grasslands, Savanna and some Azonal vegetation types).[25] As this information is collated, the assessment of habitat loss should be expanded to include aspects of ecosystem condition/degradation using similar indices.[18]

The Ecosystem Area Index is a more sensitive indicator of habitat loss at the biome and national scales than the 'remaining extent' and 'rate of loss' metrics. This is illustrated in the Desert biome where, overall, the rate of loss was only 0.02% per year between 1990 and 2018, but nearly all of the habitat loss was focused in a single small vegetation type – Alexander Bay Coastal Duneveld. As each vegetation type contributes equally to the Index for the Desert Biome, this indicator reflected a 0.15% per year decline over the same period, highlighting the threat to this vegetation type.

The extent and rate of habitat loss, and EAI are simple and powerful indicators of the state of biodiversity for national and international reporting processes.[2,18] However well-intentioned and ambitious the national or global targets are, the consequences of not meeting such targets are minimal. For example, our results indicate that South Africa failed to halve the rate of habitat loss between 2010 and 2020, which was the goal of Aichi Biodiversity Target 5. While the post-2020 successors to the Aichi targets are currently being debated by the Convention on Biological Diversity, it is clear that existing land-use planning and regulatory tools need to be strongly enforced and improved in order to change the trajectory of continued habitat loss. For example, the habitat loss metrics discussed above provide the basis for ecosystem risk assessment such as the IUCN Red List of Ecosystems[14], which allows for the identification of threatened ecosystems. These ecosystems are then prioritised in systematic biodiversity plans that underpin bioregional planning processes.[26] Threatened ecosystems are also directly referred to in South African environmental authorisation regulations[8,27], although their influence on land-use decision-making processes, especially at the local government level, needs to be strengthened.

While national and provincial authorities have taken steps recently to develop online environmental screening tools (which include key biodiversity information) and publish regulations making the use of the tools mandatory for all environmental authorisation processes, the efficacy of the interventions are difficult to gauge.[28] Monitoring the status of species and ecosystems with appropriate and up-to-date indicators improves biodiversity assessments and plans, and provides an avenue through which we can test the impact of policy interventions and regulations.[2]

Using land-cover change data to monitor ecosystems and biodiversity has its limitations, but the simplicity of the approach is also a great strength. Habitat loss related indicators are relatively easy to update (i.e. land-cover products from semi-automatic remote-sensing platforms are becoming a reality globally[9]), and the trends observed are relatively simple to interpret and are clearly attributable to drivers of change. The suite of conservation tools that use these metrics are impressive, although they do require consistent and strong implementation programmes. As we enter the UN Decade of ecosystem restoration (2021–2031) we need to complement the habitat loss indicators with analogous tools that allow semi-automated tracking of the more subtle elements of ecosystem degradation, including invasive species, overgrazing and altered fire regimes.[5]

Acknowledgements

We thank the members of the National Biodiversity Assessment – Terrestrial Reference Group and staff of the South African National Biodiversity Institute who discussed and commented on early versions of this work between 2016 and 2018. We also thank Prof. Timm Hoffman for commenting on an early draft. J.A.S. acknowledges funding from the National Research Foundation of South Africa (grant no. 118593). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests

We declare that there are no competing interests.

Authors' contributions

A.L.S.: Conceptualised the study; developed the methods; curated the data; undertook the analyses; produced the figures and maps; wrote the article; and edited the article. D.J.: Conceptualised the study and edited the article. J.A.S.: Conceptualised the study; wrote the article; revised the article; and edited the article.

References

1. Barger NN, Gardner TA, Sankaran M, Belnap J, Broadhurst L, Brochier V, et al. Chapter 3: Direct and indirect drivers of land degradation and restoration. In: Montanarella L, Scholes R, Brainich A, editors. The IPBES assessment report on land degradation and restoration. Bonn: Secretariat of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services; 2018. p. 198–314. Available from: https://www.ipbes.net/system/tdf/ipbes_6_inf_1_rev.1_2.pdf?file=1&type=node&id=16514
2. Skowno AL, Poole CJ, Raimondo DC, Sink KJ, Van Deventer H, Van Niekerk L, et al. National Biodiversity Assessment 2018: The status of South Africa's ecosystems and biodiversity – synthesis report. Pretoria: South African Naitonal Biodiversity Institute; 2019. http://opus.sanbi.org/handle/20.500.12143/6362
3. Murray NJ, Keith DA, Bland LM, Ferrari R, Lyons MB, Lucas R, et al. The role of satellite remote sensing in structured ecosystem risk assessments. Sci Total Environ. 2018;619–620:249–257. https://doi.org/10.1016/j.scitotenv.2017.11.034
4. Jantz SM, Pintea L, Nackoney J, Hansen MC. Landsat ETM+ and SRTM data provide near real-time monitoring of Chimpanzee (*Pan troglodytes*) habitats in Africa. Remote Sens. 2016;8(5), Art. #427. https://doi.org/10.3390/rs8050427

5. Slingsby JA, Moncrieff GR, Wilson AM. Near-real time forecasting and change detection for an open ecosystem with complex natural dynamics. ISPRS J Photogramm Remote Sens. 2020;166(May):15–25. https://doi.org/10.1016/j.isprsjprs.2020.05.017

6. Li BV, Hughes AC, Jenkins CN, Ocampo-Peñuela N, Pimm SL. Remotely sensed data informs Red List evaluations and conservation priorities in Southeast Asia. PLoS ONE. 2016;11(8):1–18. https://doi.org/10.1371/journal.pone.0160566

7. Skowno AL, Raimondo DC, Poole CJ, Fizzotti B, Slingsby JA, editors. National Biodiversity Assessment 2018 technical report volume 1: Terrestrial realm. Pretoria: South African National Biodiversity Institute; 2019. http://opus.sanbi.org/handle/20.500.12143/6370

8. Botts E, Skowno A, Driver A, Holness S, Maze K, Smith T, et al. More than just a (red) list: Over a decade of using South Africa's threatened ecosystems in policy and practice. Biol Conserv. 2020;246, Art. #108559. https://doi.org/10.1016/j.biocon.2020.108559

9. Gómez C, White JC, Wulder MA. Optical remotely sensed time series data for land cover classification: A review. ISPRS J Photogramm Remote Sens. 2016;116:55–72. https://doi.org/10.1016/j.isprsjprs.2016.03.008

10. GeoTerraImage. Technical report: 1990 South African national land cover dataset version 5.2. Pretoria: South African Department of Environmental Affairs; 2015. Available from: http://egis.environment.gov.za

11. GeoTerraImage. Technical report: 2013/2014 South African national land cover dataset version 5. Pretoria: South African Department of Environmental Affairs; 2015. Available from: http://egis.environment.gov.za

12. GeoTerraImage. Technical report: 2018 South African national land cover dataset version 4. Pretoria: South African Department of Environment Forestry and Fisheries; 2019. Available from: http://egis.environment.gov.za

13. Mucina L, Rutherford MC, editors. The vegetation of South Africa, Lesotho and Swaziland. Pretoria: South African Naitonal Biodiversity Institute; 2006.

14. Keith DA, Rodríguez JP, Rodríguez-Clark KM, Nicholson E, Aapala K, Alonso A, et al. Scientific foundations for an IUCN Red List of ecosystems. PLoS ONE. 2013;8(5), e62111. https://doi.org/10.1371/journal.pone.0062111

15. Jewitt D, Goodman PS, Erasmus BFN, O'Connor TG, Witkowski ETF. Systematic land-cover change in KwaZulu-Natal, South Africa: Implications for biodiversity. S Afr J Sci. 2015;111(9–10), Art. #2015-0019. https://doi.org/10.17159/sajs.2015/20150019

16. Dayaram A, Harris LR, Grobler BA, Van der Merwe S, Rebelo AG, Powrie LW, et al. Vegetation map of South Africa, Lesotho and Swaziland 2018: A description of changes since 2006. Bothalia. 2019;49(1), Art. #2452. https://doi.org/10.4102/abc.v49i1.2452

17. Scholes RJ, Biggs R. A biodiversity intactness index. Nature. 2005;434(7029):45–49. https://doi.org/10.1038/nature03289

18. Rowland JA, Bland LM, Keith DA, Juffe-Bignoli D, Burgman MA, Ferrer-Paris JR, et al. Ecosystem indices to support global biodiversity conservation. Conserv Lett. 2020;13(1), e12680. https://doi.org/10.1111/conl.12680

19. Berliner D, Desmet P. Eastern Cape biodiversity conservation plan technical report. Pretoria: Department of Water Affairs; 2007. Available from: http://dev.biodiversityadvisor.sanbi.org/ECBCP/ECBCP_technicalreport2007.pdf

20. McLean CT, Ground LE, Boon RGC, Roberts DC, Govender N, McInnes A. Durban's Systematic Conservation Assessment. Durban: eThekweni Municipality; 2016. Available from: http://www.durban.gov.za/City_Services/development_planning_management/environmental_planning_climate_protection/Publications/Documents/DurbansSystematicConservation.pdf

21. Escott B, Elliot F, Livingstone T. Draft KwaZulu-Natal biodiversity spatial planning terms and processes: Version 3.3. Pietermatizburg: Ezemvelo KZN Wildlife; 2016.

22. Pooley S. Burning Table Mountain: An environmental history of fire on the Cape peninsula. New York: Palgrave Macmillan; 2014. https://doi.org/10.1057/9781137415448

23. National Forests Act of South Africa . Notice No. 1388 South Africa: Government Gazette 19408; 1998. Available from: www.gov.za/sites/default/files/gcis_document/201409/a84-98.pdf

24. Prince S, Von Maltitz G, Zhang F, Byrne K, Driscoll C, Eshel G, et al. Chapter 4: Status and trends of land degradation and restoration and associated changes in biodiversity and ecosystem fundtions. In: Montanarella L, Scholes R, Brainich A, editors. The IPBES assessment report on land degradation and restoration. Bonn: Secretariat of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem services; 2018. p. 315–426. Available from: https://www.ipbes.net/system/tdf/ipbes_6_inf_1_rev.1_2.pdf?file=1&type=node&id=16514

25. Slingsby JA, Moncrieff GR, Rogers AJ, February EC. Altered ignition catchments threaten a hyperdiverse fire-dependent ecosystem. Glob Chang Biol. 2020;26(2):616–628. https://doi.org/10.1111/gcb.14861

26. Botts EA, Pence G, Holness S, Sink K, Skowno A, Driver A, et al. Practical actions for applied systematic conservation planning. Conserv Biol. 2019;33(6):1235–1246. https://doi.org/10.1111/cobi.13321

27. Bland LM, Nicholson E, Miller RM, Andrade A, Carré A, Etter A, et al. Impacts of the IUCN Red List of Ecosystems on conservation policy and practice. Conserv Lett. 2019;12(5), e12666. https://doi.org/10.1111/conl.12666

28. South African Department of Environment, Forestry and Fisheries. National Environmental Screening Tool [webpage on the Internet]. No date [cited 2020 Aug 04]. Available from: https://screening.environment.gov.za

Check for updates

Teatime in Kruger: Tailoring the application of the Tea Bag Index approach to an African savanna

AUTHORS:
Loewan L. Erasmus[1]
Helga van Coller[1]
Frances Siebert[1]

AFFILIATION:
[1]Unit for Environmental Sciences and Management, North-West University, Potchefstroom, South Africa

CORRESPONDENCE TO:
Loewan Erasmus

EMAIL:
loewan_erasmus@outlook.com

DATES:
Received: 23 Aug. 2019
Revised: 28 Sep. 2020
Accepted: 28 Sep. 2020
Published: 29 Jan. 2021

HOW TO CITE:
Erasmus LL, Van Coller H, Siebert F. Teatime in Kruger: Tailoring the application of the Tea Bag Index approach to an African savanna. S Afr J Sci. 2021;117(1/2), Art. #6846. https://doi.org/10.17159/sajs.2021/6846

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:
carbon cycling, herbivory, herbivore loss, nutrient-rich, semi-arid

FUNDING:
None

Attempts to obtain standardised decomposition data to determine potential drivers of carbon release have evolved from the use of cotton strips and standardised leaf litter mixtures to the most recent Tea Bag Index (TBI). The TBI is an internationally standardised method to collect comparable, globally distributed data on decomposition rate and litter stabilisation, using commercially available tea bags as standardised test kits. As this index was developed as a citizen science project in the northern hemisphere, we aimed to highlight the potential value – and pitfalls – of its application in a subtropical African savanna. We furthermore aimed to expand on existing protocol details and propose amendments to achieve an enhanced understanding of decomposition dynamics across temporal and spatial scales in African ecosystems. Proposed adaptations include extended incubation periods for long-term monitoring studies, the burial of more tea bags to account for potential losses, and the use of additional equipment to enhance effective sampling. These adaptations provide a system-specific protocol which can facilitate studies aimed to understand the interactions between top-down drivers (e.g. herbivory, fire, climate variability) and bottom-up controls (e.g. decomposition) in carbon flux dynamics of savanna ecosystems. Application of the proposed extended protocol in a semi-arid savanna provided results which reinforce the potential value of the TBI in an African context.

Significance:

- The TBI is a relatively easy and cost-effective approach to gather globally distributed data on potential decomposition rate and inherent carbon flux, yet it was developed and primarily tested in boreal and temperate ecosystems.
- The use of more paired tea bag replicates and additional equipment is a viable means to mitigate tea bag losses to several savanna-based agents of disturbance, while enabling confident conclusions made from statistical results and improved estimates of the TBI. High recovery success across disturbance treatments and incubation periods suggest that the TBI can be applied successfully to spatial and temporal decomposition studies.

Introduction

Research on linkages between decomposition and carbon flux is covered extensively for temperate[1,2] and boreal[3,4] ecosystems, whilst studies in tropical and subtropical ecosystems, especially in Africa, remain limited[5,6]. The paucity of information available on these mechanistic relationships in African savannas may be ascribed to complex interactions between top-down (e.g. climate variability, fire, large mammalian herbivores) and bottom-up controls (e.g. soil physical-chemical properties, soil-based microbes and detritivores) that collectively regulate savanna structure and function.[6,7] While governed by various environmental factors, decomposition is primarily regulated by the chemical composition of site-specific leaf litter. Cross-site comparison of natural litter decomposition and intrinsic drivers thereof is subsequently compromised by variability in detrital chemistry.[1,3,4,6,8-10]

Attempts to standardise approaches to examine the role of environmental drivers on decomposition have evolved from the use of cotton strips or natural leaf litter mixtures[1,6,9] to the most recent Tea Bag Index (TBI) approach introduced by Keuskamp et al.[8] This approach is an internationally standardised, user-friendly method to collect comparable, globally distributed decomposition data. The TBI consists of two parameters describing decomposition rate (k) and stabilisation factor (S).[5,8] Standardised test kits include commercially available Lipton® tea bags (i.e. rooibos and green tea).[8] Tea bags were selected based upon differences in chemical composition, with green tea representing high-quality organic matter with low C:N ratios, and rooibos tea a low-quality organic matter with high C:N ratios.[3,8,9] Tea bags therefore act as proxies for labile and recalcitrant compounds in naturally occurring organic matter.[1,8,9]

Assessments of ecosystem functioning, such as the TBI, are becoming increasingly valuable, especially for application in monitoring programmes in areas exposed to increasing intensities of land-use and/or climate change. However, the physical and biological template upon which African systems function varies from most ecosystems in which the TBI has been applied.[1-4] Here, we reflect on applying the TBI in an African context to facilitate decomposition monitoring. Details on (1) an experimental design aimed at elucidating temporal dynamics, (2) tea bag losses due to large mammalian herbivores, fungal infestation and termites, (3) recommended equipment to enhance effective sampling and (4) the potential value in African systems are provided.

Study area

The study was conducted in a semi-arid subtropical savanna at the Nkuhlu exclosures long-term monitoring site (24°58′S, 31°46′E) located in the Kruger National Park, South Africa. This area is characterised by a hot growing season with sporadic precipitation from October to April and a mild to warm, dry non-growing season.[11] Average annual rainfall is 561 mm, with temperatures varying from an average minimum of 5.6 °C in winter to an average maximum of 32.6 °C in summer.[11] Situated on the foot slopes of undulating granitic landscapes, the study site is characterised by sodium-rich, deep duplex soil, referred to as the sodic zone. The plant community of the sodic zone is described as a *Sporobolus nitens-Euclea divinorum* Dry Sodic Savanna and is associated with nutrient-rich vegetation.[11] Sodic patches therefore produce palatable, high-quality forage capable of supporting large herbivores, including grazers and mixed-feeders.[11]

Materials and methods

Experimental design

The Nkuhlu exclosures form part of a large-scale, long-term exclusion experiment of ecosystem drivers (i.e. herbivory and fire) to determine their effect on spatial and temporal heterogeneity patterns of vegetation in a semi-arid African savanna.[11] These exclosures consist of three herbivore treatments including (1) a partially fenced area (elephant exclosure), (2) an unfenced area (control) and (3) a fully fenced area (large mammalian herbivore exclosure).[11] As the TBI was developed and applied in systems in which decomposition is dominated by microbial agents[6,12], it does not account for the presence and possible damage caused by savanna-based disturbance agents, such as large mammalian herbivores and termites[6,10,12]. We applied a paired tea bag design which entailed the burial of 20 green and 20 rooibos tea bags in each of the 25 fixed plots across three herbivore treatments (Figure 1) which added up to 1000 experimental tea bags in total.

Decomposition is a time-bound process consisting of different phases[3,9], primarily regulated by the chemical composition of detrital substrate with respect to labile and recalcitrant ratios[5,8]. Due to the dynamic nature of savannas, regulating factors change over time, influencing the rate and extent of decomposition. Temporal comparisons are therefore essential to observe not only progression in decomposition phases, but the influence of site-specific factors thereon.[6,7,9,12,13] Our suggested extended application includes replicates which represent different incubation periods (i.e. 3-, 6-, 9- and 12-months) to capture dynamics in *k* and *S* but also to determine the resilience of the TBI in an unexplored system over the course of a year (Figure 1). All tea bags were buried in January 2019 (summer) and retrieved in intervals of three months, each representing a different season (i.e. April 2019, July 2019, October 2019 and January 2020).

Statistical analyses

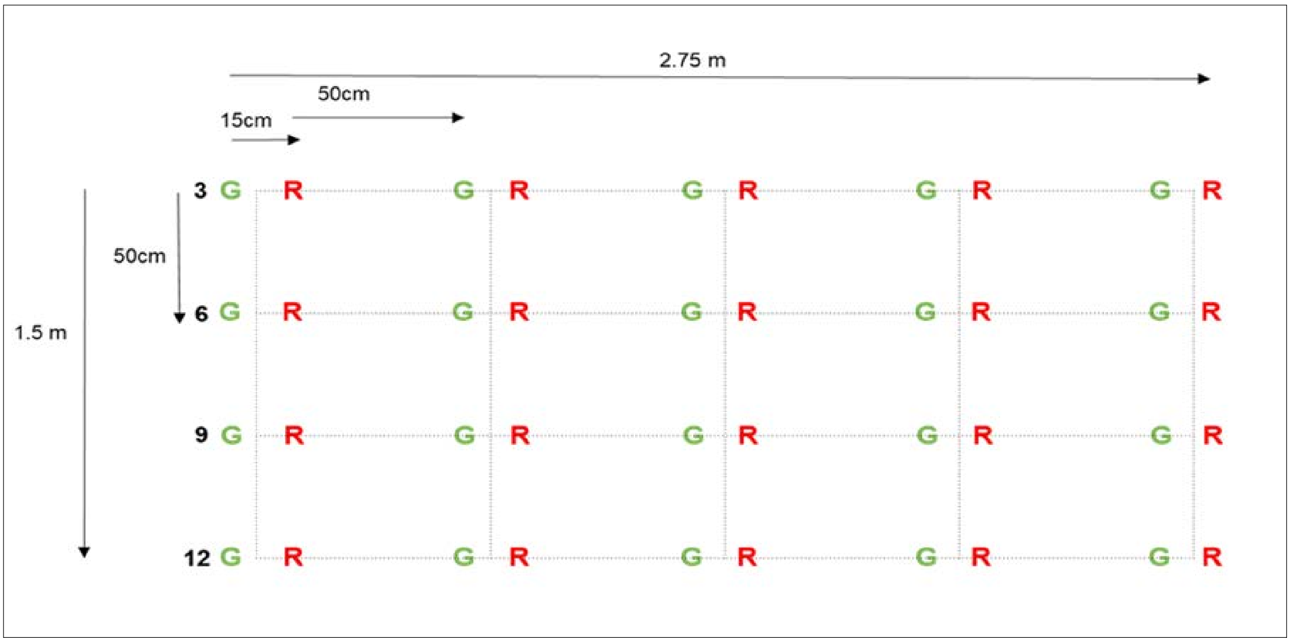
Differences in recovery success of green and rooibos tea bags within each incubation period were tested for significance using independent Student's *t* tests in Microsoft Excel (2013). To test for significant variation in the number of tea bags infested by fungi and damaged by termites across treatments and over the different incubation periods, a one-way analysis of variance (ANOVA) was applied in Paleontological Statistics Software (PAST) and Microsoft Excel (2013).

Results and discussion

The overall recovery success, irrespective of herbivore treatment and incubation, was 78.5%. Results did not reveal significant differences in recovery between green and rooibos tea among most treatments (Figure 2), although significantly fewer rooibos tea bags could be retrieved after the 6-month incubation period (Figure 2).

A finely grained black fungus was present on 42% of recovered green tea bags, while no rooibos tea bags were infested. Fungal infestation showed no significant difference between incubation periods ($p=0.897$). Fungal infestation and damage by termites were significantly lower in the control site ($p<0.001$, Figure 3a and $p=0.003$, Figure 3b). Rooibos tea bags were significantly more damaged by termites than were green tea bags ($p=0.033$).

From our results, it is evident that the presence of large mammalian herbivores complicates the retrieval of tea bags. This could be ascribed to difficulty in finding markers, because the use of above-ground markers is not recommended in areas with high mammal activity. Moreover, detritivore activity in the form of termites has a significant impact on the longevity of tea bags. To account for termite-based damage, Teo et al.[6] suggested the use of physical and/or chemical barriers (e.g. metal mesh and termiticide), although such applications may impede the practicality and standardisation of the TBI and furthermore reduce its effectiveness in measuring decomposition by altering regulatory constituents[12]. The seemingly high number of replicates used in our small-scale investigation successfully buffered the tea bag losses ascribed to animal, insect and fungal activity. Sufficient recovery (78%) with little differences between green and rooibos tea bags will lead to confident conclusions made from estimates of the TBI.



G, green tea bag; R, rooibos tea bag

Figure 1: Experimental plot layout consisting of a paired tea bag design across four incubation periods.

Table 1: Application of the original Tea Bag Index (Keuskamp et al.[8]) and extended version with potential pitfalls and recommendations for African studies

Preparations		
Keuskamp et al.[8]	**Potential pitfalls for African study**	**Recommendations**
Order tea bags (Lipton® green and rooibos) – Note that woven nylon mesh tea bag production halted in 2017 and was replaced with non-woven polypropylene tea bags. Measure initial weight of tea bag and subtract weight of empty tea bag. Mark tea bag labels using permanent marker.	Lipton tea bags need to be shipped from the Netherlands, so logistic delays should be expected. Marked labels may deteriorate when buried. Dry organic material can absorb moisture, altering the initial recorded weight.	Order non-woven polypropylene tea bags (Green -EAN 87 22700 05552 5 / EAN 8714100770542 & Rooibos -EAN 87 22700 18843 8) at least 3 months prior to the onset of the experiment. Cover marked labels with transparent adhesive tape to preserve the markings. Store weighed tea bags at room temperature in airtight containers. Mark cupcake holders and brown paper bags (i.e. used for weighing and storage of retrieved bags) in accordance with individual tea bags.

Equipment		
Keuskamp et al.[8]	**Potential pitfalls for African study**	**Recommendations**
Permanent marker Small spade Above-ground markers (e.g. stick or plastic marker)	Site-specific conditions (i.e. animal activity, compacted soil and high herbaceous biomass) require specialised equipment to minimise tea bag losses.	Transparent adhesive tape Airtight plastic containers GPS Metal detector Rope quadrant (conforming to experimental plot layout) Flat-head shovel Custom-made auger with 8-cm marker (height = 11 cm; diameter = 4.5 cm) Hammer Metal washers (galvanised steel; 1.5 cm diameter) – 1 per tea bag Resealable plastic bags (3.5 x 2.5 cm) – 1 per metal washer Brown paper bags (18 x 8.5 cm) – 1 per tea bag Cupcake holders (standard size – 5.5 cm bottom diameter) –1 per tea bag Sieve (45 cm diameter; 3.5 mm mesh) Portable carrier (field equipment and tea bags) Pharmaceutical tablet holders (10 mL)

Experimental design		
Keuskamp et al.[8]	**Potential pitfalls for African study**	**Recommendations**
Green and rooibos tea bags buried pairwise (15 cm apart) at 8 cm depth. Between 5 and 32 pairs per site (1 m apart). Recommended incubation period of 90 days.	Potential animal and termite activity might lead to tea bag losses, reducing the effectivity of few replicates. Decomposition processes and savanna systems are dynamic, necessitating extended study periods.	1000 non-woven polypropylene tea bags (500 green and 500 rooibos) buried pairwise (15 cm apart) in five plots (1.5 x 2.75 m) per site. Five sites/treatments were sampled in this study. Plots followed a grid design (four rows, 50 cm apart), representing different incubation periods in 3-month intervals. Rows contained five pairs of tea bags (five columns) buried 50 cm apart at 8 cm depth.

Burial		
Keuskamp et al.[8]	**Potential pitfalls for African study**	**Recommendations**
Bury tea bags pairwise (15 cm apart) at a depth of 8 cm. Ensure labels are visible above the soil and mark the area.	Obstructions (e.g. vegetation, termite mounds or gullies) might impede exact replicates. Soil conditions (e.g. compaction) could make burial of tea bags difficult. The study area is subjected to animal activity, therefore the potential effects this might have on the burial and recovery of tea bags should be considered. Locating plots for retrieval might be difficult.	Use a rope quadrant to produce exact replicates of plots. A custom-made mini auger can be used to create holes to bury and retrieve tea bags in hard soil, also limiting disturbance to surrounding soil. Metal washers in resealable plastic bags and label should be placed directly above tea bags 2 cm below the surface to facilitate location with a metal detector. Completely fill holes after burial to ensure tea bags are not exposed and thus susceptible to removal by animals or abiotic factors such as water run-off or wind. Take GPS coordinates at each plot.

[Keuskamp et al.[8] provide very little detail on the removal, analysis and storage of a high number of replicates.]

Retrieval		
Extended procedure for African study	**Potential pitfalls**	**Recommendations**
Locate plots with GPS coordinates and tea bags with a metal detector. Rope quadrant can be used to determine exact position of tea bags. Remove topsoil with a flat-head shovel to expose the label and/or metal washer. Place mini auger directly above the label and screw into the soil up until the 8-cm mark by striking it with a hammer. Remove tea bag from mini auger and place in corresponding cupcake holder and brown paper bag.	GPS accuracy (5 m radius) impedes exact point location. Disturbance to remaining rows (i.e. incubation periods). High numbers of retrieved tea bags may result in confusing numbers and samples. Ambient temperature and humidity may influence results.	Take a picture of the plot with the rope quadrant present. This might help identify significant structures and ease the location of metal washers. Work from the top of the plot (i.e. first row) down so as to not step on the other incubation rows. Brown paper bags absorb moisture and ensure that the tea bag remains dry and does not attract further fungal and bacterial colonisation.

Sample analysis		
Extended procedure for African study	**Potential pitfalls**	**Recommendations**
Place individual tea bags on a sieve to remove excess soil particles and roots. Place cupcake holders and corresponding tea bags in an oven and dry at 70 °C for 48 h. Weigh empty cupcake holder (to three decimal places) and remember to tare the scale. Cut open the tea bag and pour organic material into the cupcake holder and weigh (to three decimal places). Record weight on the TBI 2.0 NW data sheet **for non-woven** tea bags.	The durability of tea bags decreases with incubation and they can easily tear. Foreign material inside tea bag. Ensure that data are recorded on the correct data sheet as there are two different formats (i.e. one for woven nylon tea bags and one for non-woven tea bags).	Gently remove excess particles and debris from tea bags. Due to insect activity, mostly termites, some bags might have holes (resulting in the loss of material) and others might have foreign material inside which will affect the final weight. These tea bags should not be included in the TBI data sheet as it will affect results. As far as possible, one person should weigh all tea bags to ensure consistency not only in weighing but also in identifying these 'altered' tea bags. Visit www.teatime4science.org to download the correct data sheet.

Storage
After weighing the organic matter, the sides of the cupcake holder containing the organic matter should be folded and placed in a correspondingly labelled pharmaceutical tablet holder and stored at room temperature for potential use in the future.

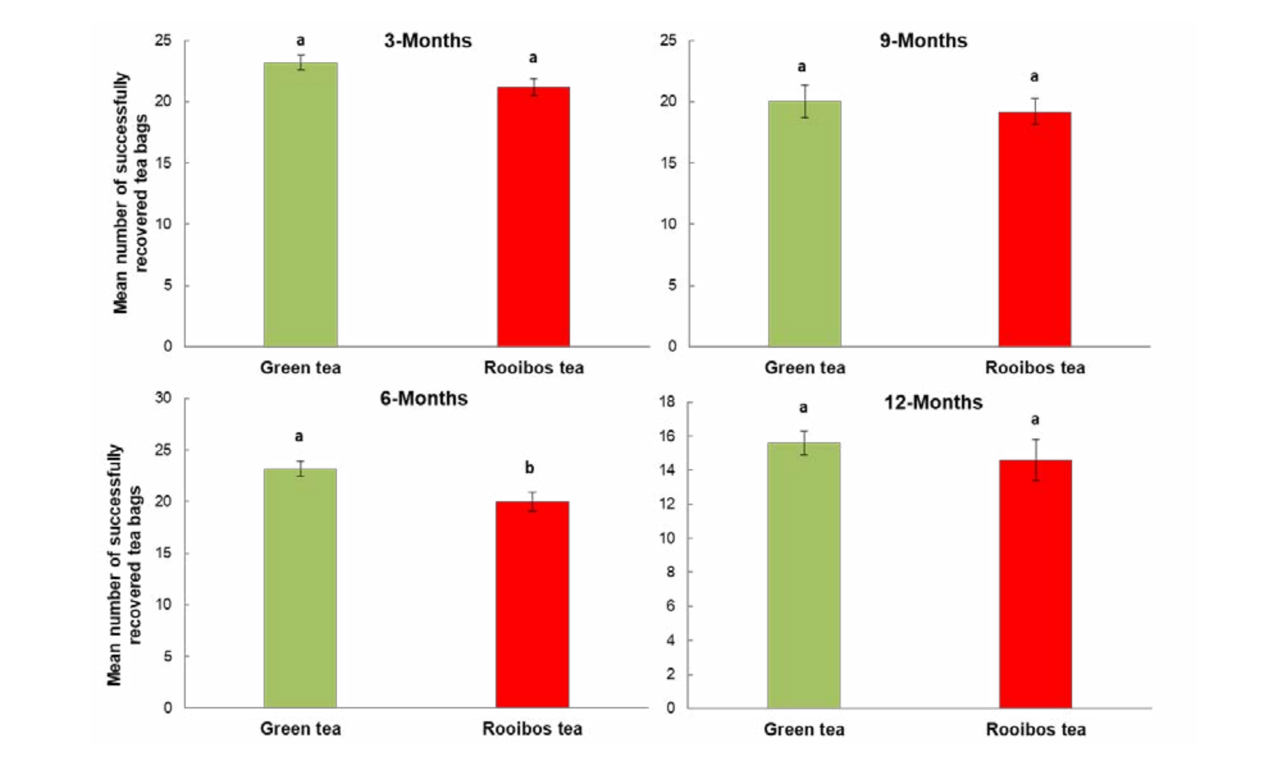
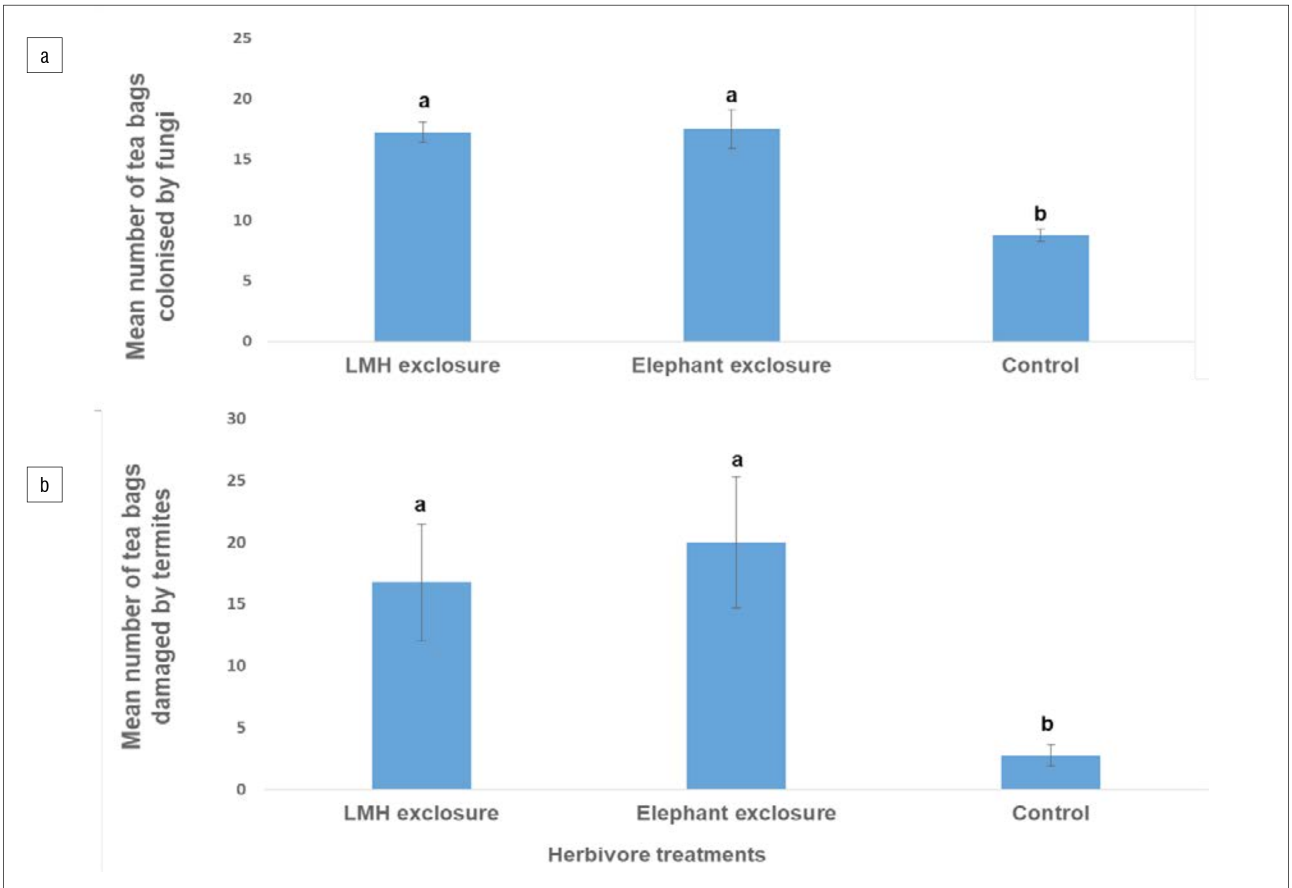


Figure 2: Mean number of successfully recovered tea bags over four incubation periods. Significant differences ($p<0.05$) between tea bags are indicated with different lowercase letters.



LMH exclosure, all large mammalian herbivores excluded; Elephant exclosure, elephants (and giraffes) excluded; Control, all herbivores present

Figure 3: Mean number of tea bags (a) infected by fungi and (b) damaged by termites across herbivore treatments. Significant differences ($p<0.05$) between herbivore treatments are indicated by different lowercase letters.

Value of the TBI in an African context

Decomposition is central to the effective functioning of terrestrial ecosystems as it forms the link between above- and belowground nutrient cycling.[3,4,6,9,14,15] Yet, little is known about site-specific disturbance effects (i.e. herbivory and fire) on decomposition and carbon flux. TBI-based decomposition studies can provide valuable information on the extent of these disturbances and their potential effect on essential ecosystem processes.[1,5,9,14,15] The TBI approach has been developed to remove the subjectivity involved in using site- and species-specific litter, only providing information pertaining to the potential decomposition capacity (i.e. deduced from standardised litter) of a specific site or system based on process-driven soil functions.[1,8,9] However, despite the major differences in chemical composition of standardised (e.g. tea) and local litter, both respond in similar ways to environmental drivers.[1,9] The TBI is therefore able to identify and examine environmental drivers of decomposition without ambiguous effects of site-specific litter, serving as a reference which facilitates data comparison across spatial scales.[1,3,5,8,9]

Despite growing concerns about climate change and increased atmospheric CO_2 levels, research on decomposition and soil carbon flux remains limited for Africa.[5,10] Necessary resources required for such studies are often unavailable in many African countries. However, citizen science projects, such as the TBI, have become a useful tool to facilitate ecological research due to relatively easy application and public engagement (e.g. Teatime 4 Science, http://www.teatime4science.org/).[1,5,8,9]

With the aim to promote efficient decomposition studies, we therefore suggest the application of this extended TBI approach for at least a 9-month duration together with application-based amendments as set out in Table 1. The TBI remains a relatively easy and cost-effective approach which greatly reduces the amount of data to be collected in normal litterbag decomposition studies. As such, this method is more conducive to both the available resources and prevailing environmental conditions in Africa.

Acknowledgements

We acknowledge and thank South African National Parks (SANParks) for general logistic support. We express our gratitude towards various people who contributed to this project through assisting in the burial and retrieval of the tea bags: Jaco Koch, Clarissa Minnaar, Marlize Muller, Judith Botha and Franco de Kock; as well as Thomas Rikombe for protection during field work.

Competing interests

We declare that there are no competing interests.

Authors' contributions

L.L.E.: Methodology, data collection, sample analysis, data analysis, validation, data curation, writing – the initial draft, writing – revisions, project leadership. H.v.C.: Methodology, data collection, sample analysis, data analysis, validation, data curation, writing – the initial draft, writing – revisions, project leadership, student supervision, project management. F.S.: Conceptualisation, writing – the initial draft, writing – revisions, project leadership.

References

1. MacDonald E, Brummell ME, Bieniada A, Elliott J, Engering A, Gauthier TL, et al. Using the Tea Bag Index to characterize decomposition rates in restored peatlands. Boreal Environ Res. 2018;23:221–235.
2. Tremblay JA, Boulanger Y, Cyr D, Taylor AR, Price DT, St-Laurent MH. Harvesting interacts with climate change to affect future habitat quality of a focal species in eastern Canada's boreal forest. PLoS ONE. 2018;13(2):1–26. https://doi.org/10.1371/journal.pone.0191645
3. Elumeeva TG, Onipchenko VG, Akhmetzhanova AA, Makarov MI, Keuskamp JA. Stabilization versus decomposition in alpine ecosystems of the Northwestern Caucasus: The results of a tea bag burial experiment. J Mt Sci. 2018;15(8):1633–1641. https://doi.org/10.1007/s11629-018-4960-z
4. Petraglia A, Cacciatori C, Chelli S, Fenu G, Calderisi G, Gargano D, et al. Litter decomposition: Effects of temperature driven by soil moisture and vegetation type. Plant Soil. 2019;435(1–2):187–200. https://doi.org/10.1007/s11104-018-3889-x
5. Becker J, Kuzyakov Y. Seasonal variation in standardized litter decomposition and effects of elevation and land use at Mount Kilimanjaro. Paper presented at: 19th EGU General Assembly; 2017 April 23–28; Vienna, Austria. p.14877.
6. Teo A, Kristensen NP, Keuskamp JA, Evans TA, Foo M, Chisholm RA. Validation and extension of the Tea Bag Index to collect decomposition data from termite-rich ecosystems. Pedobiologia. 2020;80, Art. #150639. https://doi.org/10.1016/j.pedobi.2020.150639
7. Ratnam J, Sheth C, Sankaran M. African and Asian savannas: Comparisons of vegetation composition and drivers of vegetation structure and function. In: Scogings PF, Sankaran M. Savanna woody plants and large herbivores. Chichester: John Wiley & Sons Ltd.; 2019. p. 25–49. https://doi.org/10.1002/9781119081111.ch2
8. Keuskamp JA, Dingemans BJ, Lehtinen T, Sarneel JM, Hefting MM. Tea Bag Index: A novel approach to collect uniform decomposition data across ecosystems. Methods Ecol Evol. 2013;4(11):1070–1075. https://doi.org/10.1111/2041-210X.12097
9. Didion M, Repo A, Liski J, Forsius M, Bierbaumer M, Djukic I. Towards harmonizing leaf litter decomposition studies using standard tea bags – a field study and model application. J For. 2016;7(8):167. https://doi.org/10.3390/f7080167
10. Leitner M, Davies AB, Parr CL, Eggleton P, Robertson MP. Woody encroachment slows decomposition and termite activity in an African savanna. Glob Chang Biol. 2018;24(6):2597–2606. https://doi.org/10.1111/gcb.14118
11. Van Coller H, Siebert F, Siebert SJ. Herbaceous species diversity patterns across various treatments of herbivory and fire along the sodic zone of the Nkuhlu exclosures, Kruger National Park. Koedoe. 2013;55(1), Art. #a1112. https://doi.org/10.4102/koedoe.v55i1.1112
12. Sundsdal A. Teatime for termites: Understanding the impact of human land-use and climate on termite and microbial litter decomposition [master's thesis]. Trondheim: Norwegian University of Science and Technology; 2019.
13. Berg B, Kjønaas OJ, Johansson MB, Erhagen B, Åkerblom S. Late stage pine litter decomposition: Relationship to litter N, Mn, and acid unhydrolyzable residue (AUR) concentrations and climatic factors. Forest Ecol Manag. 2015;358:41–47. https://doi.org/10.1016/j.foreco.2015.08.032
14. Eldridge DJ, Delgado-Baquerizo M, Travers SK, Val J, Oliver I, Hamonts K, et al. Competition drives the response of soil microbial diversity to increased grazing by vertebrate herbivores. Ecology. 2017;98(7):1922–1931. https://doi.org/10.1002/ecy.1879
15. Paudel E, Dossa GG, de Blécourt M, Beckschäfer P, Xu J, Harrison RD. Quantifying the factors affecting leaf litter decomposition across a tropical forest disturbance gradient. Ecosphere. 2015;6(12):1–20. https://doi.org/10.1890/ES15-00112.1

Check for updates

AUTHOR:
Robin Fisher[1]

AFFILIATION:
[1]College of Life and Environmental Sciences, University of Exeter, Penryn, Cornwall, United Kingdom

CORRESPONDENCE TO:
Robin Fisher

EMAIL:
rcvfisher2016@gmail.com

DATES:
Received: 29 Mar. 2020
Revised: 16 Sep. 2020
Accepted: 17 Sep. 2020
Published: 29 Jan. 2021

HOW TO CITE:
Fisher R. Possible causes of a substantial decline in sightings in South Africa of an ecologically important apex predator, the white shark. S Afr J Sci. 2021;117(1/2), Art. #8101. https://doi.org/10.17159/sajs.2021/8101

ARTICLE INCLUDES:
☒ Peer review
☐ Supplementary material

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

EDITOR:
Bettine van Vuuren

KEYWORDS:
marine predators, marine conservation, bycatch, *Carcharadon carcharias,* KwaZulu-Natal, Gansbaai

FUNDING:
None

Possible causes of a substantial decline in sightings in South Africa of an ecologically important apex predator, the white shark

A decline in sightings of a top predator, the white shark (*Carcharadon carcharias*), in South Africa was quantified in order to identify possible causes for this decline. White shark sightings data across 8 years (2011–2018), recorded from a cage-diving vessel in Gansbaai are reported. A significant decline in mean total white shark sightings per boat trip (>6 in 2011 to <1 in 2018) and a 69% reduction in the probability of a sighting were found. Correlating with this decline in sightings is a rise in sightings of sevengill sharks (*Notorynchus cepedianus)* in False Bay and copper sharks (*Carcharhinus brachyurus)* in Gansbaai, as well as substantial ecosystem changes. The effects of lethal conservation measures such as the use of shark nets in KwaZulu-Natal; the direct and indirect effects of overfishing including a reduction in smoothhound (*Mustelus mustelus)* and soupfin (*Galeorhinus galeus*) sharks; and novel predation on white sharks are discussed as possible causative factors for this decline in white shark sightings.

Significance:

- The results of this paper highlight the need to reassess the impact of marine conservation initiatives and fishing practices. Failure to do so could seriously affect ecologically and economically important marine species. This paper reveals a potentially serious decline to the South African white shark population, characterised by a substantial decline in white shark sightings. This decline correlates with the overfishing of prey species, bycatch, the use of lethal gill nets and ecological changes such as the novel presence of orca. Better marine management is required if South Africa wishes to keep a healthy white shark population.

Introduction

Of the more than 1000 species of sharks and rays, approximately half are classified by the International Union for Conservation of Nature (IUCN) as vulnerable, endangered, critically endangered or near threatened.[1] Between 63 million and 273 million sharks, belonging to 61 species, are killed by human activity each year – mainly from legal and illegal fishing operations.[2] The white shark, *Carcharadon carcharias*, is an example of one of these species.[3] White sharks perform important regulatory roles in coastal ecosystems by influencing the structure and function of communities via both direct and indirect predatory effects.[4] Their removal has been shown to cause reductions in marine biodiversity and cause negative trophic cascades.[5] For instance, white sharks control the population of meso-consumers, like Cape fur seals (*Arctocephalus pusillus)*.[6] Without white shark predation, and a reduction in the threat of their predation, the fur seal foraging range has extended, resulting in the removal of refugia for seal prey species[7] which has reduced the populations of economically significant fishery species.[4-6]

White sharks also provide fiscal benefits to the South African economy.[8] For example, the Gansbaai shark cage-diving industry raised USD4.4 million in 2003 and brings an estimated USD2 million into False Bay annually, excluding multiplier effects (e.g. hospitality industry).[9] More recently in Gansbaai, a complete white shark absence for 21 days was predicted to have caused a loss of ZAR1.5–2 million in January 2016.[10] Therefore white sharks are considerably important to South Africa's economy.

Historically, white sharks were exploited by fisheries and are still caught as bycatch.[11] They are also still legally culled despite being granted protected status in South Africa in 1991.[11] For example, from 1976 to 2008, 1073 sharks were killed in shark nets in KwaZulu-Natal.[11] This resulted in a 99% reduction in adult white sharks caught.[11,12] There is also evidence to suggest that the white shark population is declining.[6] For example, at Seal Island in False Bay, an overall non-significant decline in white shark sightings occurred up to 2014 and sightings were the lowest on record during 2016–2018.[9] Given the economic importance of white sharks to the tourism industry and their wider ecological significance, identifying the reasons for this decline in sightings is vital. Data on sightings from 2011 to 2018 in Gansbaai, a white shark hotspot on the Western Cape, were used to determine some of the possible causes of the decline in white sharks in South Africa.[6]

Methods

The data used for this study were collected between January 2011 and December 2018 from an ongoing survey of sightings. Data collection occurred in all months except November 2011, October 2012, January 2016 and June 2018. In total, 92 months of data were analysed. The annual trend in sightings across the 8 years was analysed. Overall, 3010 trips were undertaken and 15 122 shark sightings were documented. In this context, one sighting equates to at least one shark seen within the duration of a trip, and is not the total number of individuals sighted within the trip. Shark sightings per year varied from a peak of 2644 in 2013, and a minimum of 360 in 2018. To control for sampling effort, the total number of shark sightings was divided by the total number of trips (Table 1).

Table 1: The mean(±s.d.) number of white shark sightings per trip per year and the total number of trips

Year	Mean shark sightings per trip	Total number of trips
2011	5.50±3.66	297
2012	7.31±3.63	345
2013	7.21±1.66	412
2014	5.84±2.69	457
2015	4.39±2.76	341
2016	4.73±2.72	424
2017	1.21±1.73	386
2018	0.81±1.04	348

Observations were recorded 9 km southeast of Gansbaai, in the proximity of Dyer Island and Geyser Rock (collectively known as 'shark alley') and at an inshore reef system known as Joubertsdam. A 10.6-m ecotourism cage-diving boat, operated by White Shark Projects, collected all the data used in this project. Trips were conducted up to three times a day, weather permitting. The trips started at 07:00, 10:00 and/or 13:00 and lasted 3 h. Sharks were attracted to the vessel via an olfactory cue of chum consisting of local fish farm discards and by throwing and dragging a tuna head attached to a rope in the water. These cues were provided for the entire duration of every trip.

All statistical analyses were conducted using the statistical software 'R studio 2018'.[13] The distribution of the shark sightings per trip data was bimodal, because some trips had zero sightings. Therefore, a binomial generalised linear model was used to test if the probability of sighting a shark changed per year. This model was tested using an analysis of variance (ANOVA), with a chi-squared test. Then, shark sightings ≥1 were normally distributed, so a linear model was used to investigate if shark sightings per trip ≥1 differed per year. An ANOVA was used to test the significance of this model.

Results

The total reduction in the probability of sighting a shark per trip was 69% over the 8-year period from a maximum of 100±0.0% in 2013 to a minimum of 31±0.48% in 2018 ($X^2_{1,177}$=135.6; p<0.0001). From 2011, the mean number of shark sightings decreased by 0.81 sightings per trip each year ($F_{1,141}$=14.8; p<0.001) (Figure 1). In total, over eight years, mean white shark sightings per trip declined by 6.5. Initially, mean white shark sightings increased to a peak in 2012 of 7.31±3.63. However, in 2013, a gradual decline in mean white shark sightings per trip began. With the exception of 2016, on average, each year saw a decline of 1.3 white shark sightings per trip. At the end of this study period in 2018, white shark sightings per trip were at a minimum of 0.81±1.04.

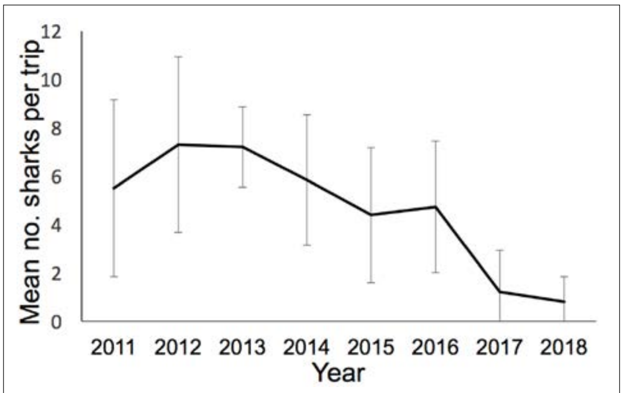


Figure 1: The trend in mean (±s.e.) number of shark sightings per trip, across the 8-year project.

Discussion

The main finding of this research project was that white shark sightings substantially decreased from >6 sharks per trip to <1 shark per trip over the 8-year study period, in a region internationally recognised for its high white shark abundance.[6] A recent project assessing population size using mark-recapture and genetic sampling between 2009 and 2013 concluded that it was reasonable to assume white sharks sampled at Gansbaai were representative of the whole South African population.[6] The aforementioned project estimated that the white shark population consisted of between 353 and 522 individuals, with a small contemporary effective population size of 333.[6] Therefore, if the sightings data in this paper are indicative of a decline in the white shark population, then this population is at serious risk of extinction.

It is likely that direct mortality of white sharks has contributed to the decline in white shark sightings seen here. For example, the KwaZulu-Natal Sharks Board reported that an average of 16.8 white sharks were killed in their nets per year between 2013 and 2017.[14] This number equates to ~84 white sharks killed in total in 5 years.[14] Assuming Andreotti et al.'s[6] population estimate of the South African white shark was accurate, then shark nets were responsible for between ~one-sixth to a ~one-third decline in the white shark population from 2013 to 2017.[6,14] Moreover, white sharks had a mean spatial overlap with fishing effort of 64% (median 65%) in the southwest Indian Ocean and have been shown to swim in the highest-risk zone in all oceans in which they were tracked.[15] In fact, since 2016, two white sharks were reported as bycatch by the South African demersal shark longline (DSL) fishery – one fatally in May 2019.[16] It is likely that the true number of white shark bycatch in South Africa is underreported by the DSL fishery.[17] For example, in 2017, the South African Department of Agriculture, Forestry and Fisheries ran a bycatch assessment experiment from a (DSL) vessel in Algoa Bay.[18] In just one day of operation, three white sharks were caught[18] (Fallows C 2020, written communication, July 25). Moreover, in Australia, white shark bycatch is reported to be six times higher than that in South Africa, with a reported estimated catch of >30 white sharks per year in 2014.[19] Therefore it is likely white sharks are subjected to a high risk of becoming bycatch from longline fishing[16], and might still be killed at unsustainably high levels in KwaZulu-Natal shark nets[14].

Overfishing of white shark prey species could also be a significant contributor to the decline in white sharks in South Africa.[20] Analyses of the stomach contents of 225 white sharks caught between 1978 and 2009[21] and 591 white sharks from 1974 to 1988[22] caught in the KwaZulu-Natal nets, indicated that for ~75% of their lifespan, elasmobranch and teleost prey made up ~60% of their diet. Dusky sharks (*Carcharhinus obscurus)* and small sharks including smoothhound (*Mustelus mustelus)* and soupfin (*Galeorhinus galeus)* sharks were the main prey items.[22] From 2013 to 2017, dusky sharks were also the most commonly caught shark in KwaZulu-Natal nets, with an annual average mortality of 78.6 individuals per year.[14] Due to overfishing, substantial declines in smoothound[23] and soupfin sharks also correlated with the decline in white shark numbers[24]. Following a reallocation of the fishery in 2013, and a fishing effort increase in 2015, the soupfin shark population dropped to ~13% of carrying capacity[23], and the smoothhound population decreased by 30%[24]. Around this time, fishing pressure was strong in False Bay but has now shifted eastwards to near Mossel Bay.[24] Furthermore fishing effort has increased[16] – for example, smoothhound have been fished by the DSL at an extent 1.6–2.2 times higher than that recommended for all South African fisheries, including 17 588 (~123 t) individuals in 2016 and an estimated 23 592 (~165 t) individuals in 2019.[16]

Overfishing has also led to a reduction in pelagic sharks including shortfin mako (*Isurus oxyrinchus)* and blue (*Prionace glauca)* sharks.[25] These sharks are usually preyed upon by a pelagic ecotype of orca (*Orcinus orca)*.[25] In 2017, this type of orca was observed inshore and five liverless white shark carcasses characteristic of orca predation washed up in Gansbaai.[26] Therefore, overfishing may have severely decreased prey stocks, causing orcas to predate on white sharks.[20] In response, white sharks may have temporarily left the Western Cape waters to evade the new predation pressure.[20] A similar response by white sharks to orca predation was documented at the Farallon Islands, USA.[27] This could help explain the sudden drop in white shark sightings from 2017 onwards, but it is likely to have a medium-term effect.[20] White shark sightings started declining in 2013, prior to the arrival of orca.[20] Therefore, whilst significant, orcas cannot fully explain the decrease in white shark sightings.

The decline in white shark sightings also correlated with substantial changes in trophic structure.[19] Notably sevengill sharks (*Notorynchus cepedianus)*[19], and copper sharks (*Carcharhinus brachyurus)* have become prevalent in False Bay and Gansbaai[28]. Furthermore, possibly due to a reduction in white shark predation, the foraging range of the Cape fur seal has extended.[16] This extension has caused a removal of refugia for endangered African penguins (*Spheniscus demersus)* and their prey species.[7,29] Consequently, African penguins must extend their foraging ranges, making them more susceptible to predation.[29] These changes also correlated with a further reduction in bony fish species and may have deleterious consequences on the distribution of kelp forests.[16] Therefore, prioritising white shark conservation could mitigate against the loss of other endangered species, biodiversity and economically important fishery species.

Conclusion

There has been a clear reduction in white shark sightings in Gansbaai – an area that was internationally recognised as having the highest white shark abundance worldwide.[6] More research is needed to quantitatively and qualitatively assess the ecological significance of this decline in white sharks. However, if sightings of white sharks continue to decrease, then there is likely to be a negative effect on the distribution of other coastal species, including a reduction in the biodiversity of South Africa's marine ecosystem.[16] The extent to which white sharks have temporarily disappeared or possibly relocated is still uncertain.[20] Therefore long-term population surveys and monitoring of white shark sightings around the whole South African coastline is required.[15] Yet, despite these uncertainties, there is strong evidence highlighting the negative impacts of a poorly managed DSL fishery.[16] This fishery might be directly and indirectly impacting the white shark population. Given the ecological and economic importance of the white shark to South Africa, and the likelihood that their population has declined severely, better fisheries management is imperative. In addition, there is evidence indicating that the use of shark nets is detrimental to the population of white sharks.[14] Therefore, a re-evaluation of the use of shark nets should be considered to prevent the complete loss of the white shark from South Africa.

Acknowledgements

I thank Tom Slough and Georgina Vermuelen at White Shark Projects for their hard work and support during the writing of my dissertation and for giving me access to White Shark Project's data set, on which this article is based. I also thank the University of Exeter, and my supervisor Professor Andy Russell for his guidance and support whilst writing my dissertation, on which this article is based. The article also benefitted substantially from comments made by the anonymous reviewers.

Competing interests

To the best of my knowledge, there are no conflicts of interest to declare.

References

1. Bland LM, Keith DA, Miller RM, Murray NJ, Rodríguez JP. Guidelines for the application of IUCN Red List of Ecosystems Categories and Criteria. Version 1.1. Gland: IUCN; 2017. https://doi.org/10.2305/IUCN.CH.2016.RLE.3.en
2. Worm B, Davis B, Kettemer L, Ward-Paige CA, Chapman D, Heithaus MR, et al. Global catches, exploitation rates, and rebuilding options for sharks. Mar Policy. 2013;40:194–204. https://doi.org/10.1016/j.marpol.2012.12.034
3. Kock A, O'Riain MJ, Mauff K, Meÿer M, Kotze D, Griffiths C. Residency, habitat use and sexual segregation of white sharks, *Carcharodon carcharias* in False Bay, South Africa. PLoS ONE. 2013;8(1), e55048. https://doi.org/10.1371/journal.pone.0055048
4. Heithaus MR, Frid A, Wirsing AJ, Worm B. Predicting ecological consequences of marine top predator declines. Trends Ecol Evol. 2008;23(4):202–210. https://doi.org/10.1016/j.tree.2008.01.003
5. Ferretti F, Worm B, Britten G, Heithaus M, Lotze H. Patterns and ecosystem consequences of shark declines in the ocean. Ecol Lett. 2010;13(8):1055–1071. https://doi.org/10.1111/j.1461-0248.2010.01489.x
6. Andreotti S, Rutzen M, Walt SVD, Heyden SVD, Henriques R, Meÿer M, et al. An integrated mark-recapture and genetic approach to estimate the population size of white sharks in South Africa. Mar Ecol Prog Ser. 2016;552:241–253. https://doi.org/10.3354/meps11744
7. Wcisel M, O'Riain MJ, Vos AD, Chivell W. The role of refugia in reducing predation risk for Cape fur seals by white sharks. Behav Ecol Sociobiol. 2014;69(1):127–138. https://doi.org/10.1007/s00265-014-1825-5
8. Gallagher AJ, Huveneers CP. Emerging challenges to shark-diving tourism. Mar Policy. 2018;96:9–12. https://doi.org/10.1016/j.marpol.2018.07.009
9. Pfaff MC, Logston RC, Raemaekers SJPN, Hermes JC, Blamey LK, Cawthra HC, et al. A synthesis of three decades of socio-ecological change in False Bay, South Africa: Setting the scene for multidisciplinary research and management. Elem Sci Anth. 2019;7, Art. #32. https://doi.org/10.1525/elementa.367
10. Mabaleka NM. The contribution of shark cage diving tourism to coastal economies: A case study of a coastal town in the Western Cape, South Africa [doctoral dissertation]. Cape Town: Cape Peninsula University of Technology; 2020.
11. Dulvy NK, Fowler SL, Musick JA, Cavanagh RD, Kyne PM, Harrison LR, et al. Extinction risk and conservation of the world's sharks and rays. eLife. 2014;3, Art. #00590. https://doi.org/10.7554/elife.00590
12. Peschak T. Shark nets – the real killers of the sea. Afr Geogr. 2009;17:36–44.
13. R Core Team. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2018. Available from: https://www.R-project.org/
14. Shark Catch Statistics [document on the Internet]. c2017 [cited 2020 Jul 30]. Available from: https://www.shark.co.za/Pages/SharkCatchStats
15. Queiroz N, Humphries NE, Couto A, Vedor M, Da Costa I, Sequeira AM, et al. Global spatial risk assessment of sharks under the footprint of fisheries. Nature. 2019;572(7770):461–466. https://doi.org/10.1038/s41586-019-1444-4
16. Creecy B. Parliamentary report back on the DSL [webpage on the Internet]. c2019 [cited 2020 Jul 30]. Available from: https://sharkfreechips.com/facts/south-african-government-ignoring-facts-scientists-fisherman/
17. Da Silva C, Booth A, Dudley S, Kerwath S, Lamberth S, Leslie R, et al. The current status and management of South Africa's chondrichthyan fisheries. Afr J Mar Sci. 2015;37(2):233–248. https://doi.org/10.2989/1814232x.2015.1044471
18. DSL Vessels are catching Cites protected species and are fishing in MPAs [webpage on the Internet]. c2020 [cited 2020 Jul 30]. Available from: https://sharkfreechips.com/facts/dsl-vessels-are-catching-cites-protected-species-and-are-fishing-illegally-in-mpas/
19. Taylor SM, Braccini JM, Bruce BD, McAuley RB. Reconstructing Western Australian white shark (*Carcharodon carcharias*) catches based on interviews with fishers. Mar Freshw Res. 2018;69(3):366–375. https://doi.org/10.1071/mf17140
20. Hammerschlag N, Williams L, Fallows M, Fallows C. Disappearance of white sharks leads to the novel emergence of an allopatric apex predator, the sevengill shark. Sci Rep. 2019;9(1), Art. #1908. https://doi.org/10.1038/s41598-018-37576-6
21. Cliff G, Dudley SF, Davis B. Sharks caught in the protective gill nets off Natal, South Africa. 2. The great white shark *Carcharodon carcharias* (Linnaeus). Afr J Mar Sci. 1989;8(1):131–144. https://doi.org/10.2989/02577618909504556
22. Hussey NE, McCann HM, Cliff G, Dudley SF, Wintner SP, Fisk AT. Size-based analysis of diet and trophic position of the white shark (*Carcharodon carcharias*) in South African waters. In: Domeier ML, editor. Global perspectives on the biology and life history of the white shark. Boca Raton, FL: CRC Press; 2012. p. 27–49.
23. Da Silva C, Winker H, Parker D, Kerwath S. Assessment of smoothhound shark *Mustelus mustelus* in South Africa [article on the Internet]. c2019 [cited 2020 Jul 30]. Available from: https://www.researchgate.net/profile/Henning_Winker/publication/338491221_Assessment_of_smoothhound_shark_Mustelus_mustelus_in_South_Africa/links/5e1782ee4585159aa4c2d6f9/Assessment-of-smoothhound-shark-Mustelus-mustelus-in-South-Africa.pdf

24. Winker H, Parker D, Da Silva C, Kerwath S. First comprehensive assessment of soupfin shark *Galeorhinus galeus* in South Africa [article on the Internet]. c2019 [cited 2020 Jul 30]. Available from: https://www.researchgate.net/profile/Henning_Winker/publication/338491033_First_comprehensive_assessment_of_soupfin_shark_Galeorhinus_galeus_in_South_Africa/links/5e17817b4585159aa4c2d6b1/First-comprehensive-assessment-of-soupfin-shark-Galeorhinus-galeus-in-South-Africa.pdf

25. Williams A, Petersen S, Goren M, Watkins B. Sightings of killer whales *Orcinus orca* from longline vessels in South African waters, and consideration of the regional conservation status. Afr J Mar Sci. 2009;31(1):81–86. https://doi.org/10.2989/ajms.2009.31.1.7.778

26. Engelbrecht TM, Kock AA, O'Riain MJ. Running scared: When predators become prey. Ecosphere. 2019;10(1), e02531. https://doi.org/10.1002/ecs2.2531

27. Pyle P, Schramm MJ, Keiper C, Anderson SD. Predation on a white shark (*Carcharodon carcharias*) by a killer whale (*Orcinus orca*) and a possible case of competitive displacement. Mar Mamm Sci. 1999;15(2):563–568. https://doi.org/10.1111/j.1748-7692.1999.tb00822.x

28. McKay T. Locating great white shark tourism in Gansbaai, South Africa within the global shark tourism economy. In: Rogerson JM, Visser G, editors. New directions in South African tourism geographies. Cham: Springer International Publishing; 2020. p. 283–297. https://doi.org/10.1007/978-3-030-29377-2

29. Vanstreels RE, Parsons NJ, McGeorge C, Hurtado R, Ludynia K, Waller L, et al. Identification of land predators of African penguins *Spheniscus demersus* through post-mortem examination. Ostrich. 2019;90(4):359–372. https://doi.org/10.2989/00306525.2019.1697971