South African
Journal *of* Science

volume 117
*number 11/12*

Using indigenous wood from southern Africa as tonewood

Determination of alcohols in hand sanitisers

A validated method for the analysis of nyaope

Life cycle assessment of single-use and reusable plastic bottles

Gender representation in scientific organisations

Innovative building through 3D printing

ASSAf
ACADEMY OF SCIENCE OF SOUTH AFRICA

Volume 117
Number 11/12
November/December 2021


eISSN: 1996-7489

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

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

**Correspondence and enquiries**
sajs@assaf.org.za



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

## Review Article



## Research Articles



**Cover caption**

Two African violins (photo: Martina Meincken). In an article on page 137, Meincken and colleagues show that several southern African wood species are suitable to be used as tonewoods and that the sound produced with such a violin is – although somewhat different – of high quality, comparable to the sound quality of violins made traditionally from spruce and maple wood. The violin on the left was made from yellowwood and sapele (by Hannes Jacobs) and the violin on the right was made from blackwood, hardpear and sapele (in the Department of Forest and Wood Science at Stellenbosch University with the help of Hadley Dumini).

# A journal's dilemma

Some years ago, I was working with a graduate student who was interested to research issues of transformation in a South African institution. The student was an insider to the institution and was aware of many of the strides the institution had made towards change but also knew the many challenges and disappointments along the way. I met with colleagues of the student, who were enthusiastic about the study. All that remained in terms of access was the go-ahead from the non-executive chair of the board of the organisation.

When the student and I met the chair of the organisation, the chair informed us that she was in support of the study in principle, but she could not approve the study as the methods were so bad. The student had chosen to use qualitative methods with an ethnographic component; according to the board chair, these methods were inappropriate as they were not scientific and the qualitative sampling would lead to skewed results. The board chair (who was not a researcher, and, as far as I know, had never completed any graduate research training) informed me that only quantitative methods were scientific and that proper sampling was required. My own suspicion about this response was that the concerns were not actually methodological but, instead, based on a fear of reputational damage to the organisation as a result of the research. I was proved right in this suspicion when we presented the board chair with a revised proposal which comprised a substantial quantitative component and a commitment to survey every single person who was part of the organisation. At this point, the board chair said that any research done on the organisation could cause damage, and that she would not allow us to proceed. The student had to abandon this particular project.

There are many issues in this story which are relevant to considering the functioning of science in the public domain. Some, for example, have to do with the problems of the permeable boundary in community-based research between, on the one hand, community representatives who, as far as is possible, speak on behalf of stakeholder groups, and, on the other, gatekeepers who may block access for a range of reasons not necessarily fully aligned with a careful consideration of what the best interests of a community are.[1] For the purposes of this discussion, though, I want to point to the conflation between methodological concerns and criticisms as first expressed by the board chair, and other concerns which were not initially expressed. The question of method here seems to have been used as an acceptable smokescreen for worries about possible reputational damage to the organisation.

As a science journal, our commitment is to publish work which, to borrow terms from our mission and vision statements, is 'excellent' and 'high quality'. Traditionally, and in our journal, the assessment of the quality of research articles depends heavily on methodological considerations, as our reviewer guidelines reflect.

But we have a further commitment, and this is a commitment to publish, as we say on our website, 'for the benefit of scholars, educators, the general public and policymakers'. We also promise to provide 'a forum for discussion of news and developments in research and higher education'. These criteria are not methodological in the narrow sense, but ideological. We commit ourselves to exercise judgement about whether research is 'of benefit'. This means we can decide, within our remit, not to publish work which may be methodologically and theoretically sound in a narrow sense but is also, in our view, for example, racist, sexist, xenophobic, homophobic, or disablist. We make no apology for this commitment, especially in light of the many past shameful abuses of science to oppressive and even murderous ends.[2] We also believe that discussion and debate is important for the development of science and the academy more broadly – through dissent and argument, ideas develop and grow.

All of this looks quite simple, but there is an implicit contradiction. As a journal, and especially as a journal funded by the public purse in South Africa, we strive to promote views which are in line with our view of what may be best for society. But it would be arrogant of us to suggest that we are arbiters of the public good, simply by virtue of our being a science journal. Indeed, one of the values we espouse as scientists is that debate is to be encouraged as a necessary part of the development of ideas. But we cannot have proper debate unless we publish views which are not fully aligned with our own. We may have our own beliefs about when authors are wrong, mischievous, or even destructive in their submissions, but we are not censors. We are aware of how the concept of 'free speech' has been abused by right-wing activists and how difficult and entangled discussions of 'free speech' can be[3,4] – just google the words 'your right to say it' to find distasteful examples. So these issues are not simple or uncontested.

As a journal, as are other journals, we are committed to what we believe is good for science and society – a part of which is the tolerance of and engagement with dissent. This faces us with the oft-cited dilemma of giving platform for views with which we disagree, or may find abhorrent.

We do not have a solution to the dilemma we face – we cannot hide fully behind the argument that our decisions are made on the basis of an incontrovertible standard of what constitutes sound methods and argument. Nor can we escape the challenge that values play in making our decisions. We are aware that we will have readers who question the motives of some of our authors and who are angered by our decision to publish material with which they (and we) may disagree. But our current approach to this is to err on the side of allowing discussion, and encouraging readers to engage in debate. There may well be those who will ask how we came to publish certain contributions as they see them to be not in the public interest. We encourage these interlocutors to submit their own views and debates for publication. We appeal to all our readers to help us navigate these complex issues.

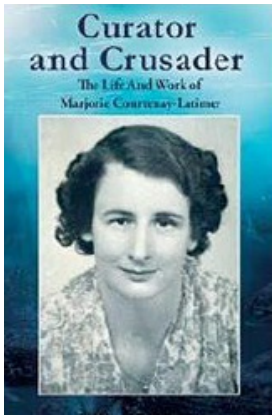
## References


1. Fregonese F. Community involvement in biomedical research conducted in the global health context; what can be done to make it really matter? BMC Med Ethics. 2018;19(suppl 1), Art. #44. https://doi.org/10.1186/s12910-018-0283-4
2. Véliz C. Three things digital ethics can learn from medical ethics. Nat Electron. 2019;2(8):316–318. https://doi.org/10.1038/s41928-019-0294-2
3. Titley G. Is free speech racist? Cambridge, UK: Polity; 2020.
4. Ahmed A. Your right to say it: Good and bad faith in the free speech debate. Times Literary Supplement. 2021;6179:26–27.



**HOW TO CITE:**
Swartz L. A journal's dilemma. S Afr J Sci. 2021;117(11/12), Art. #12772. https://doi.org/10.17159/sajs.2021/12772

**BOOK TITLE:**
Curator and crusader: The life and work of Marjorie Courtenay-Latimer




**AUTHOR:**
Mike Bruton

**ISBN:**
9780639932699 (softcover, 292 pp)

**PUBLISHER:**
Footprint Press, Hermanus, South Africa; ZAR250

**PUBLISHED:**
2019

**REVIEWER:**
Mary Cole[1,2]

**AFFILIATIONS:**
[1]East London Museum, East London, South Africa
[2]Department of Zoology and Entomology, Rhodes University, Makhanda, South Africa

**EMAIL:**
marybursey@elmuseum.za.org

**HOW TO CITE:**
Cole M. The extraordinary life of a pioneering museum curator. S Afr J Sci. 2021;117(11/12) Art. #12075. https://doi.org/10.17159/sajs.2021/12075

**ARTICLE INCLUDES:**
☐ Peer review
☐ Supplementary material

**PUBLISHED:**
29 November 2021

# The extraordinary life of a pioneering museum curator

Most people with an interest in nature have surely heard of Marjorie Courtenay-Latimer in connection with the discovery of the first living coelacanth known to science. Her efforts put East London and its new museum on the world map. I joined the staff of the museum in 1988, so I knew 'Marge' during her 80s and 90s when she often visited the museum to be filmed about the coelacanth or to attend museum functions. Reading *Curator and Crusader*, I enjoyed learning so much more about this remarkable woman and what she achieved during her 42 years as curator and subsequent years of active retirement. Mike Bruton remembers Marge from his school days and gives her credit for the positive impact she had on his career. He realised that she 'lived a story worth telling' and did an excellent job piecing together this highly recommended biography from diverse sources of information, greatly assisted by Nancy Tietz, ex-director of the East London Museum (ELM). He used transcripts of Marge's father's diaries, an unpublished manuscript on the history of her family, Marge's illustrated scrapbooks, copious autobiographical notes and minutes of meetings, letters, speeches, newspaper clippings and field notes in the ELM archives. Extracts are often reproduced verbatim – a style I found informative, and which allows readers to make up their own minds about the content. There are ample illustrations: black and white photos, sketches and 16 pages of colour photos. The six appendices include the full citation for the award of her honorary doctorate and titles of a selection of her newspaper articles.

Among my favourite chapters are the two on her childhood, giving insights into how her keen interest in the world around her developed, encouraged by her mother. At the age of nine she was collecting and sending specimens to the museum in King Williams Town! Her father changed jobs or was transferred every year or so to remote railway stations or villages all over the Eastern Cape and Free State Provinces. A map of the whole country shows where these places are, although I would have preferred a larger map of only the relevant regions, and to see the location of the family farms – important places for the cultivation of Marge's interests. Her formal education was very rudimentary, but on the positive side she enjoyed opportunities to explore the natural world in many different environments. Although Marge excelled in many subjects, as one of six daughters, there were no finances for tertiary education. Surprisingly for someone so full of vitality and who lived to 97, Marge was a weak and sickly child, but she did survive the Spanish flu pandemic of 1918.

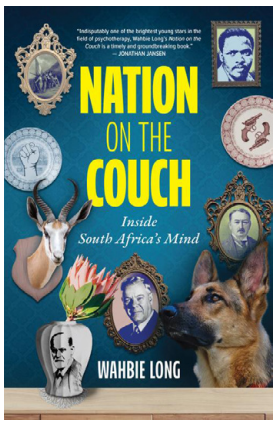
Not only was Marge disadvantaged in terms of education, she was also no stranger to other hardships such as the desperate food shortages of World War 1. She was appointed first curator of the East London Museum in 1931, a period when there were few women in leading roles. She must have experienced some tough times with the Museum Board and professional scientists given that she was a woman, young and lacked formal education. She was even ridiculed by some for suggesting that the coelacanth specimen was significant and worth saving for science. Her success is an inspiration, and shows how doors do open for those with passion and enthusiasm, who dream big and work hard.

Marge had very wide knowledge, and although she was not an academic, Bruton draws attention to her scientific research in ornithology and botany; the extensive reference list includes all her scientific articles. She communicated with a large number of leading scientists across a range of disciplines and sought the advice of experts regularly; extracts of these communications are enlightening. Evidence is presented that she was held in high regard by the majority of scientists and by her staff, although the few who did not respect her work are mentioned. The coelacanth was named *Latimeria* by JLB Smith as a tribute and he dedicated his bestselling *Old Fourlegs* to Marge. There is a chapter on her crusades for environmental conservation, and her contributions were praised by the Director of Nature Conservation, Douglas Hey. A chapter is devoted to bird-ringing as Marge used bird rings in her research 12 years before the official start of the practice in South Africa, something which is not acknowledged elsewhere.

No one can dispute that Marge excelled in her mandate and there are several chapters on how she built the ELM into a thriving hub of community activity assisted at every turn by the Chairman of the Board, GG Smith. Detail on buildings, finances and staffing are buffered by lighter elements like their annual collecting trips to remote wildernesses and pet dogs. Early in her career she was sent for training at the Durban and Cape Town museums and started forming relationships with museologists; she was a founder member of the South African Museums' Association. The very large dioramas Marge created were the first to make full use of this technique in South Africa and still delight visitors today. She was passionate about educating the public and mounted an extraordinary number of temporary exhibitions, held talks and guided tours, and was instrumental in founding the local culture- and nature-based societies. She was a prolific writer of short articles, including in a regular column in the local press. Marge loved children and always welcomed them to show her their finds. In short, her museum was her life.

Marjorie Courtenay-Latimer the woman is portrayed in the final four chapters. Personal traits and circumstances, heart-wrenching and humorous, are captured in lengthy extracts from interviews with people who had worked with her or known her well. She described her life as a happy one. I remember well her sense of humour and her giggling when telling a funny story. One of my few disappointments in the book is the lack of more anecdotes by current or recently retired staff of the ELM. She was described, not as a feminist, but as very feminine, and she always wore blouses, skirts and stockings even in the field!

**BOOK TITLE:**
Nation on the couch: Inside South Africa's mind




**AUTHOR:**
Wahbie Long

**ISBN:**
9781990973307 (softcover, 200 pp)

**PUBLISHER:**
Melinda Ferguson Books, Cape Town, South Africa; ZAR280

**PUBLISHED:**
2021

**REVIEWER:**
Crain Soudien[1]

**AFFILIATION:**
[1]School of Education, University of Cape Town, Cape Town, South Africa

**CORRESPONDENCE TO:**
Crain Soudien

**EMAIL:**
crain.soudien@uct.ac.za

**HOW TO CITE:**
Soudien C. The mind of South Africa: A review of *Nation on the Couch*. S Afr J Sci. 2021;117(11/12), Art. #12567. https://doi.org/10.17159/sajs.2021/12567

**ARTICLE INCLUDES:**
☐ Peer review
☐ Supplementary material

**PUBLISHED:**
29 November 2021

# The mind of South Africa: A review of *Nation on the Couch*

Books are meant to provoke. They are written to push, prod and stimulate. Hopefully, they help the reader to pause and to think over an idea or a thought. Amazon, shortly after it had announced the release of *Nation on the Couch*, received two customer reviews, one of which, astonishingly, pronounced that the book was a 'sad read', 'nothing new, just an endless rehashing of the same downtrodden ideas.... My only question is', asked the customer, 'Has the writer ever met a white South African?' Long has certainly provoked. That he has not succeeded in helping our dear Amazon customer to pause is certainly cause for me to pause. What will it take, I ask, because I certainly have been shaken up by the book. What will it take, please, South Africans – actually, people everywhere – for us to talk about the state of the world in which we find ourselves? What do people like our Amazon customer wish to hear? I have my suspicions but let me rather stay on the firmer ground of my opinions and make my own mistakes.

*Nation on the Couch* is an analysis, beyond the daily seductive commentary tumbling off our devices, gadgets and, still, thankfully, our newspapers, about the state of South Africa. Long describes it as 'an attempt... to excavate, as it were, the *political unconscious* of life in South Africa. It... examines both the exteriority and interiority of social life' (p. 14). The questions which prompted the writing of the book included 'why...South Africa appear(s) to lurch from one social crisis to the next? Why... her citizens brutalise each other, both interpersonally and structurally?' (p. 14). As an analysis it is focused on cataloguing and examining what he describes as 'the monsters of our collective deep' (p. 14).

It is important to describe how Long undertakes this task of analysis. It is uncompromisingly conceptual. Ideas and theories frame his argument. We will come to what those are in a moment. But the approach is also deeply grounded. He inserts himself, almost biographically, into the story he is seeking to analyse. He is present. In this, he is, of course, fulfilling and exemplifying the role and responsibility of the psychoanalyst. He does so, however, courageously, struggling, as he says of the psychoanalyst, '... to sustain an emotionally meaningful contact with one's own experience' (p. 181). That he has struggled is evident in the design of the book. It begins laden with gloom but ends, even as he is 'called out' by 'woke' intellectuals and students (p. 54), in reasoned optimism.

The conceptual heart of the text is the idea of alienation. It is, Long argues, 'the defining malady of our age' (p. 31). The book devotes three large chapters to its emotional expression: Shame, Envy and Impasse. These constitute Long's diagnosis of the 'monsters of the collective deep'. The remainder of the book, two chapters, explores the concepts of Hope and Empathy, providing what Long (p. 37) describes as 'the search for answers' to the diagnosis.

Long's understanding of alienation essentially comes from that of theorist Rahel Jaeggi who describes it as, ultimately, 'indifference and internal division, but also powerlessness and relationless with respect to oneself and to a world experienced as indifferent and alien.... the inability to establish a relation to other human beings, to things, to social institutions and thereby also ... to oneself' (p. 27). Important for Long, and so distinguishing him from much mainstream psychology, is his explanation of it as 'a bridging concept that brings into view the interconnectedness of social structure and individual experiences' (p. 32). Here interiority, made visible through Freudian analysis, is held in symbiotic connection with exteriority, explained through Marxist structural analysis. The connection is explored well in Long's insightful discussions of Shame and Envy. The opportunity is there too, perhaps less explicitly developed, in what Long describes as the state of Impasse.

The chapters on Shame and Envy are provocative. Shame, Long says, as an alienated and alienating emotion, is the complex product of the denial of respect. Denied respect and feeling undervalued, human beings 'become resentful, and if they reckon within themselves a failure to meet the standards of others they feel ashamed' (p. 40). '(W)hen social formations compromise the dignity of marginalised groups as a matter of routine, the consequences', Long argues, 'are devastating, involving either self-hating shame or envious resentment.' (p. 43). He uses this to describe and explain the catalogue of problems which have befallen South Africa, therapeutic non-compliance in the face of diseases such as HIV/Aids and tuberculosis, interpersonal violence, alcohol abuse, violent protests and many more.

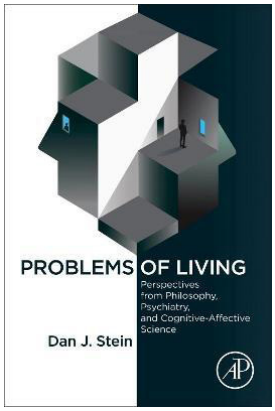
Shame is linked to Envy, Long's next major theme. Shame triggers, he argues, resentment. This resentment simmers: '[When it is] incapable of regulating itself, it finally explodes in the form of *envy* [Long's emphasis], taking shape in the seemingly wanton destruction of property and the apparent senseless resort to interpersonal violence' (p. 70). Long explains that 'in its essentials, it involves a recognition that the Other... has possession of a resource on which our tenuous existence depends. The terror that comes with the ever-present threat of annihilation is unavoidable: its retaliatory bent is all too human.' (p. 71) Long uses this envy to explain the behaviour of the students in the protests of 2015, especially those in the #RhodesMustFall movement: 'The more the Fallists destroy the institution, the more impoverished the collective ego feels. Their envy grows stronger still...' (p. 86).

Long (p. 119) evokes the 'state of Impasse' idea to explain the situation in which white South Africans find themselves in the post-apartheid period. That state he contends, drawing on Hussein Bulhan, is an essentially tragic one: 'On the one hand, they need the colonised to remain in their place... a fate that no human being will tolerate indefinitely. On the other, since the colonial relationship is effectively a recapitulation of the master–slave dialectic, the coloniser never feels recognised as human – because the act of recognition is made by a slave and is therefore worthless' (p. 119). This leads to the essential *Impasse* diagnosis where whites, uncertain of their place, 'withdraw... from public spaces into fortified enclaves' (p. 120).

After this searing analysis, the book concludes, as Long says, with a shift in register: 'from diagnosis to treatment in the search for answers' (p. 36–37). This conclusion begins with a re-siting and a recontextualising of the mood of hope, from its location in the individual to its embeddedness in material relations. To realise that mood, Long explains, is going to take work at both individual and social levels. It is about the restoration of the dignity of South Africans – recognising their mental states and, critically, implementing what he calls 'The Golden Rule', the title of his final chapter: Treating others as you wish to be treated and not as you wish *not* to be treated.

*Nation on the Couch* is a great contribution to the discussion about South Africa. Not only is it written beautifully, but it succeeds in taking complex ideas and puts them in elegantly clear terms. It will not bring *all* of us, as our Amazon customer shows, to that point where we stop and think, but it will do so for many of us. It will do so, moreover, in ways which Long has anticipated. Some of us will agree with him, but some of us not. Some will be infuriated by his depiction of our envy. We will say – and here, actually, is the rub – that he has some of the story right but not the whole of it. There is a great deal more for us to think and say about traumatic experiences such as racism and how we as human beings have responded to it.

**BOOK TITLE:**
Problems of living: Perspectives from philosophy, psychiatry and cognitive-affective science




**AUTHOR:**
Dan Stein

**ISBN:**
9780323902397 (paperback, 326 pp)
9780323904391 (ebook, 326 pp)

**PUBLISHER:**
Academic Press, Cambridge, MA, USA; USD127.50

**PUBLISHED:**
2021

**REVIEWER:**
Anton A. van Niekerk[1,2]

**AFFILIATIONS:**
[1]Philosophy Department, Stellenbosch University, Stellenbosch, South Africa
[2]Director – Centre for Applied Ethics, Stellenbosch University, Stellenbosch, South Africa

**EMAIL:**
aavn@sun.ac.za

**HOW TO CITE:**
Van Niekerk AA. Interdisciplinary perspectives on the traditional 'big' questions of philosophy. S Afr J Sci. 2021;117(11/12), Art. #12214. https://doi.org/10.17159/sajs.2021/12214

**ARTICLE INCLUDES:**
☐ Peer review
☐ Supplementary material

**PUBLISHED:**
29 November 2021

# Interdisciplinary perspectives on the traditional 'big' questions of philosophy

Dan Stein, Professor and Head of the Department of Psychiatry at the University of Cape Town, is one of South Africa's most outstanding academics. His research output is beyond prolific. He is the most cited author in South African academia; his Google *h*-index is >145. He holds doctorates in medicine, psychiatry and philosophy, and has built his department into an internationally acclaimed centre of excellence.

*Problems of Living* must be acknowledged as one of the most remarkable achievements to which his already acclaimed accolades attest. It is an excellent example of what can be achieved in terms of interdisciplinary work, when this work is the outcome, not of a dilettante, but of a true specialist in different fields – in this case psychiatry, philosophy and neuroscience.

That an author of such distinction writes a book of such erudition (308 pages, 1984 references and a voluminous array of footnotes – the latter mainly for the sake of acknowledging the possible literature that could be consulted on almost every topic that he covers) is enough to intimidate readers. Yet, that does not happen. Stein is knowingly or unknowingly a member of a growing group of leading academics (akin to the Dutch philosopher Paul van Tongeren) who write, not for the sake of displaying accomplished erudition, but primarily for the sake of producing work that is accessible to both an academic and a popular readership. Without claiming that Stein has achieved this consistently and throughout the entire book, the effort as it stands is a quite remarkable achievement.

What this reviewer appreciated more than everything else, is Stein's conviction that the traditional 'big' questions of philosophy – What is the relation between reason and passion? Is there any sense in pain and suffering? How can we distinguish between good and bad? When are we dealing with truth? What is the meaning of life? – are not only still valid, but can be approached from a variety of the interdisciplinary perspectives, and that significant (or at least notable) progress can be made with such an enterprise, without any claim that all questions have been definitively or adequately answered.

The book unfolds in nine chapters. After the introduction, which deals with general psychiatric and philosophical perspectives on raising the 'big questions', Chapter 2 deals with the enigmatic issue of the nature of and relationship (if any!) between 'brain' and 'mind'. His argumentation in this chapter brings Stein to the point where he decides that the distinction confuses more than it clarifies. He decides to move along with his newly coined notion of 'brain-mind' – a construct that accommodates both aspects without finally clarifying all the conceptual and anatomic conundrums of the original distinction.

Chapter 3 is about the relationship between reason and passion, followed by Chapter 4 on 'The pleasures of life'. Here he carefully analyses the philosophical and psychiatric dimensions of happiness and does not hesitate to push the analysis through to an effort to understand real concrete pleasures, such as food and drink, play and music, exercise and running, and physical and natural beauty. Chapter 5 deals with the perplexing matter of pain and suffering, Chapter 6 with the distinction between good and bad and the nature and challenges of morality, and Chapter 7 with the issue that formatted both classical and modern philosophy more than anything else, namely the difference between truth and falsity.

Any reader that has made it to this point, cannot but be overawed by the prospect of the second last chapter – Stein's analysis of the meaning of life itself. Space does not allow me to elaborate on the plethora of surprising and useful insights that this chapter delivers. What this reviewer appreciates the most, though, is Stein's achievement in breaking down to manageable proportions a question that many philosophers, particularly from the analytical tradition, deem unanswerable, due to both its complexity and its tendency to evaporate into subjectivity. One quotation will hopefully demonstrate the profundity of insight that, in this chapter, Stein so amply demonstrates: "One useful contrast is between approaches that focus on the understandability of life (or determining the *meaning of life*) versus approaches that focus on what makes a life worthwhile (or finding *meaning in life*)" (p. 200, Stein's italics).

This quotation is indicative of the way in which Stein diffuses the argument that reflecting on the meaning of life is a waste of time. We mostly know what we know on the basis of the analysis of issues into smaller, manageable parts, as happens in most of the sciences. The quote above, however, illustrates that philosophy does distinguish itself from the other sciences in that its questions neither coincide with the formal questions of mathematics, nor with the empirical questions of the sciences. Philosophy deals with another kind of question. The complexity of that other kind of question cannot now be fleshed out, except to insist that that question sometimes is about wholes rather than parts – wholes such as the (meaning of) the totality of life and the universe itself.

The last chapter is 'Metaphors of life'. Here the term 'life', which we also find in the title of the book, figures most prominently. We are thereby reminded of what the book in first and last instance is about, namely the investigation of a range of interdisciplinary foci from both the human and the applied sciences on the variety of actual problems that we encounter in our efforts to live the lives of ordinary people on a daily basis. Stein rightly acknowledges the significance of the phenomenon of metaphor – that remarkable cognitive disposition that, as Aristotle saw for the first time, enables us to see similarities in dissimilars, and to thereby access meaning via the juxtaposition of entities that seemingly and *prima facie* have nothing in common: 'Brevity is the soul of wit'; 'There is a tide in the affairs of men which, taken at the flood, leads on to fortune', to quote only two of these remarkable cognitive instruments provided by Shakespeare. In this last chapter, Stein investigates metaphors such as 'life as a game', 'life as a story or narrative', 'parenting as gardening', life as either a cycle or as growth, development and progression, and, inevitably, life as a journey.

Whether the journey metaphor is the most appropriate for our understanding of life, remains to be seen. There can, however, be no doubt that reading Stein's book is a journey through a masterpiece. It is one of the most enriching experiences that this reviewer has had in a long time.

**BOOK TITLE:**
Scientometrics for the humanities and social sciences


**AUTHOR:**
Radhamany Sooryamoorthy

**ISBN:**
9780367626860 (hardback, 170 pp, GBP120.00)
9780367627010 (paperback 170 pp, GBP36.99)
9781003110415 (ebook, 170 pp, GBP33.29)

**PUBLISHER:**
Routledge, London, UK

**PUBLISHED:**
2020

**REVIEWER:**
Nelius Boshoff[1,2]

**AFFILIATIONS:**
[1]Centre for Research on Evaluation, Science and Technology (CREST), Stellenbosch University, Stellenbosch, South Africa
[2]DSI–NRF CoE in Scientometrics and Science, Technology and Innovation Policy (SciSTIP), Stellenbosch University, Stellenbosch, South Africa

**EMAIL:**
scb@sun.ac.za

**HOW TO CITE:**
Boshoff N. Exploring the potential of scientometrics for the humanities and social sciences: Towards the future. S Afr J Sci. 2021;117(11/12), Art. #12478. https://doi.org/10.17159/sajs.2021/12478

**ARTICLE INCLUDES:**
☐ Peer review
☐ Supplementary material

**PUBLISHED:**
29 November 2021

# Exploring the potential of scientometrics for the humanities and social sciences: Towards the future

Professor Sooryamoorthy's inspiration for writing a book on scientometrics for the humanities and social sciences (HSS) originated with a question he personally had wrestled with as a social scientist: why do HSS scholars seem indifferent to scientometric studies in their own disciplines, while the few HSS scholars who do conduct scientometric studies mostly do so with the science disciplines in mind? Considering this question, his goal for the book was to attract HSS scholars to the field of scientometrics by being both informative and illustrative – informative, by giving researchers in the HSS a broad overview of the scientometric approach, and illustrative, by showing, through cases, the application of scientometric techniques in the HSS. Sooryamoorthy's book consists of five chapters, each of which aims to bring to the attention of HSS scholars the potential of scientometrics as a method for descriptive and evaluative analyses in the HSS.

Chapter one will appeal to anyone interested in learning about the origins of scientometrics, its link with the sociology of science, and how the emergence of scientometrics ties with the emergence of citation indexes for the study of science. This is in addition to a discussion of typical 'laws' that provide useful models for the behaviour of scientometric indicators. Exploring the differences between scientometrics and bibliometrics in this chapter was of particular interest to me, as the two terms are often used interchangeably. Perhaps the book could also be called bibliometrics for the HSS, as the focus remains largely on publication data. It is important to note that metrics other than bibliometrics can also be regarded as instances of scientometrics. Examples in the South African context are financial and human resource indicators derived from the South African National Survey of Research and Experimental Development, and from the Higher Education Management and Information System (HEMIS) of the South African Department of Higher Education and Training (DHET). Chapter two builds on the general orientation of chapter one by focusing on the current applications of scientometrics across disciplines, with special attention given to the types of analyses that are possible.

Chapter three is the first of two chapters that explicitly deal with the HSS. It highlights reasons for the limited interest in scientometrics as an analytical approach in the HSS. Among the main reasons are the divergent publication patterns of disciplines in the HSS, and the insufficient publication coverage of HSS outputs in the main databases used for scientometric analysis. Sooryamoorthy brings together the insights of various studies in this regard and emphasises solutions: the replacement or supplementation of large international databases with national databases that have complete coverage of publication outputs in the HSS (not only journal articles but also monographs and other forms of output) and the collection and analysis of data from a set of journals specific to the HSS (which may not be indexed in the international databases). These proposals also reflect the principles and approaches set out by the European Network for Research Evaluation in the Social Sciences and Humanities (ENRESSH). For instance, to improve research evaluation for the social sciences and humanities (SSH), ENRESSH encourages all stakeholders to:

> *Develop databases reflecting all types of SSH research output,* [which must be] *useful for researchers as means of dissemination and information retrieval* [and to] *reflect upon the role of national and international authoritative lists of publication channels, and the definition of minimal standards for scholarly publications.*[1]

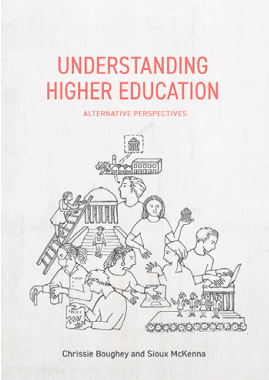
Chapter four is especially noteworthy as it gives a 'voice' to existing scientometric studies in the HSS, by systematically presenting and discussing 24 such cases. The cases span different HSS disciplines (e.g. drama, political science and tourism) and were chosen to illustrate applications such as mapping disciplines and subjects, collaboration and co-authorship, and citation analysis. What may be missing are similar cases to illustrate what Sooryamoorthy describes in chapter five as a promising way forward to attract HSS scholars to scientometrics: the analysis of qualitative data. The most important parts of qualitative information in the scientometric analysis of publications are found in the titles of publications, the publications' keywords, abstracts and cited references and obviously in the full text of publications. Although the examples of subject mapping in chapter four briefly refer to content analysis based on abstract keywords, more examples of applications of content analysis in scientometric studies (of which there are also South African examples[2,3]) would have enriched the concluding chapter, especially an explanation of the difference between quantitative and qualitative analysis of qualitative content (text) in scientometric analysis.

Sooryamoorthy has delivered a unique book to guide HSS researchers in the application of the scientometric approach, a book which also paves the way for the future. The future could be one in which HSS scholars view collections of publications as rich sources of qualitative data for scientometric analysis in their respective disciplines.

## References

1. ENRESSH. Challenges of the evaluation of social sciences and humanities research (SSH) [document on the Internet]. c2017 [cited 2021 Sep 27]. Available from: https://enressh.eu/wp-content/uploads/2017/09/Guidelines_SSH_final.pdf
2. Du Toit J, Boshoff N, Mariette N. Normative versus actual methodologies in planning research: A hybrid picture. J Plan Educ Res. 2017;37(4):477–487. https://doi.org/10.1177/0739456X16658095
3. Graham TM, Ismail T. Content and method trends in the Journal of Community Psychology between 2003 and 2007. J Community Psychol. 2011;39(2):121–135. https://doi.org/10.1002/jcop.20420

**BOOK TITLE:**
Understanding higher education: Alternative perspectives




**AUTHORS:**
Chrissie Boughey
Sioux McKenna

**ISBN:**
9781928502210 (paperback, 180 pp)
9781928502227 (free ebook, 180 pp)
9781928502234 (epub, 180 pp)

**PUBLISHER:**
African Minds, Cape Town, South Africa; ZAR250

**PUBLISHED:**
2021

**REVIEWER:**
Kathy Luckett[1]

**AFFILIATION:**
[1]Institutional Planning Department, University of Cape Town, Cape Town, South Africa

**EMAIL:**
kathy.luckett@uct.ac.za

**HOW TO CITE:**
Luckett K. *Understanding Higher Education* – Critically. S Afr J Sci. 2021;117(11/12), Art. #12444. https://doi.org/10.17159/sajs.2021/12444

**ARTICLE INCLUDES:**
☐ Peer review
☐ Supplementary material

**PUBLISHED:**
29 November 2021

# *Understanding Higher Education* – Critically

This book is an excellent critical introduction to contemporary issues facing higher education. It is based on the knowledge and experience of two authors who have worked in and researched South African higher education for over two decades, at both institutional and national levels in roles that have covered student, staff, research and institutional development as well as contributed to national policy development. The book offers substantial commentary and critical analysis on the following topics: policy and dominant policy discourses in higher education; the discursive positioning of students and the implications thereof for teaching and learning; the importance of knowledge for higher education curricula; and the impact of institutional history, culture and type on changes and challenges to the academic job and academic project, policy implementation and possibilities for transformation. The authors correctly argue that, although written from the South African higher education context, the book holds relevance for all readers interested in understanding more about higher education, especially for those from Southern contexts. The book is a must-read for higher education researchers and practitioners in a range of roles – senior managers, policymakers, academics and postgraduate students, especially those on higher education studies programmes.

A key strength of this book is the way the authors provide an accessible introduction to their deployment of Roy Bhaskar's Critical Realism (an 'underlabouring' philosophy (p.21)) and Margaret Archer's Social Realism (a sociological theory of change over time) for researching higher education with a concern to promote social justice. They are particularly concerned to 'produce a theorized response to questions about the relationship between teaching and learning in higher education, and its role in reproducing the status quo' (p.136) – and how this might be explained in order to be challenged. The authors' deployment of critical and social realism allows them to gather data from surface-level descriptions of the goings-on in higher education institutions, such as observations, experiences and empirical regularities that arise from the events that take place in higher education institutions and wider society. However, in order to explain the emergence of these data and why some things do or do not work in certain contexts, the authors demonstrate how researchers of higher education need to dig deeper to understand how phenomena emerge as an effect of the interplay between human agency and the historically sedimented structures and cultures that condition particular contexts. While not being observable, it is these 'real' mechanisms that shape or constrain individuals' potential to exercise agency and thus the extent to which change can occur.

Boughey and McKenna are skillful in the way they put Archer's social realist theory and its concepts to work in subsequent chapters. In Chapter 3 they discuss how dominant global discourses such as the shift from the idea of a welfare state to neoliberalism's 'new public management', have shaped higher education polices since World War II, in ways that promote higher education as a 'private good' and students as consumers – undermining the idea that higher education should be state-funded as a 'public good'. Turning to South African higher education as a case study, the authors critically analyse policy developments since 1994 that aimed to make the South African higher education system more equitable as well as more efficient. Drawing on social realism, the authors show how historical socio-cultural conditioning during the apartheid era continues to shape the ways different types of institution have responded to the state's new policies.

In Chapter 4, the authors challenge readers to interrogate their discursive constructions of students, and particularly the prevalent notion of the student as 'decontextualised learner'. They argue that this discourse works to position individual students as personally responsible for their academic success or failure (p.54). Consequently, learning is understood as 'a set of skills or competencies which are a-social, a-cultural and a-political' (p.71). This understanding of the student and its attendant pedagogical 'solutions' partly explain why success rates in South African higher education continue to be strongly differentiated by race. The persistence of this discourse allows universities to get away with their failure to address the effects of classism and racism that still work as causal mechanisms on South African campuses long after the political demise of apartheid: 'That the university, through its current practices, plays a role in reinforcing the unjust social status quo is a bitter pill to swallow' (p.58). According to the authors, contributing factors include 'the language problem', whereby the individual student is blamed for failing to adequately assimilate into the dominant culture and acquire the dominant colonial language (English). This misrecognises students' identities and cultural resources and works as a barrier to academic success for many students. Furthermore, acquiring the academic literacy practices valued in the disciplines is a complex process involving identity shifts for most students; this acquisition process is a lot easier for students already familiar with middle-class literacy practices. The authors sum up this excellent chapter by arguing that, as long as students are constructed as 'deficient' and 'disadvantaged', universities will fail to offer them the requisite cultural, linguistic, and pedagogical resources required for learning the distinctive knowledge-making practices of disciplinary and professional fields.

In Chapter 5, Boughey and McKenna offer a nuanced discussion on the curriculum challenges currently facing higher education institutions, especially the importance of weighing up how to respond to the legitimate interrogation of whose knowledge and interests are served by our inherited curricula against the need to provide all students with access to theoretical knowledge. In keeping with social realist theory, they show, with South African examples, how access to this 'powerful' abstract curriculum knowledge is conditioned by social and institutional histories and contexts. They argue that access to abstract knowledge is necessary if the 'public goods' of society are to be distributed equitably.

Chapter 6 provides a critical summary of the pressures facing academics and the academic project from neoliberal ideology and the global 'knowledge economy'. A consequence has been a shift in power from academics to senior management, the casualisation of academic labour, the introduction of performativity and measurable accountability checks on academics such as quality assurance and staff appraisal and, consequently, a hollowing out of collegiality and a culture of trust.

**AUTHORS:**
Margaretha Viljoen[1]
Peter F. Levay[2]

**AFFILIATIONS:**
[1]Department of Psychiatry, University of Pretoria, Pretoria, South Africa
[2]Department of Internal Medicine, Kalafong Hospital, University of Pretoria, Pretoria, South Africa

**CORRESPONDENCE TO:**
Margaretha Viljoen

**EMAIL:**
mviljoen@webafrica.org.za

**HOW TO CITE:**
Viljoen M, Levay PF. Dance as physical exercise for older people. S Afr J Sci. 2021;117(11/12), Art. #11831. https://doi.org/10.17159/sajs.2021/11831

**ARTICLE INCLUDES:**
☐ Peer review
☒ Supplementary material

**KEYWORDS:**
physical exercise, dance, older people, health

**PUBLISHED:**
29 November 2021

# Dance as physical exercise for older people

**Significance:**

- Adequate physical exercise is important for the somatic and mental health of older people.
- Dance, as an alternative to structured exercise, could provide both health and social aspects.
- The potential to adapt dance style and intensity to accommodate physical limitations makes dancing suitable for many older adults.

The importance of adequate physical activity for older people cannot be overestimated. Evidence exists that physical exercise can improve quality of life; benefit brain areas involved with executive control; delay the age-related decline in functional ability; offer a degree of protection against certain physical and mental disorders; reduce the prevalence of falls by improving flexibility, balance, and muscle strength; and potentially promote longevity.[1-3] On cellular and molecular levels, aging is said to be marked by genomic instability, telomere attrition, epigenetic alterations, impaired protein homeostasis, deregulation of systems involved in nutrient sensing, a decline in mitochondrial integrity and biogenesis, cellular senescence, stem cell exhaustion, and changes in intercellular communication.[4,5] Plausible indications are that physical activity, especially aerobic and resistance training, can have positive anti-aging effects through attenuation of such cellular and molecular mechanisms.[4]

Not everyone likes structured exercise; however, various alternatives exist. One alternative is recreational dancing – an activity primarily selected for fun and social interaction, rather than health benefits. We briefly examined the potential benefits of amateur/recreational dance, as a form of exercise, for the health and well-being of older people. A literature search was performed through electronic databases as well as manually. Electronic databases searched were EBSCOhost Research Interface databases; Academic Search Complete; APA; CINAHL and MEDLINE.

## Health benefits of dance for older people

The effects of dance on the health and general well-being of the older population (≥65 years), as appraised in 16 reviews, including 11 systematic reviews, covering a total of 296 individual studies, are summarised in Supplementary table 1. Several reviews concluded that dance has positive/protective effects on cognitive functions[6-13], such as a delay in the onset of dementia[11], on memory[7], including immediate and delayed recall[7], and on executive functions[6]. Several studies claimed dance benefits quality of life[9,12], mental states such as depression, anxiety[9,14] and general mood[12], as well as psychosocial phenomena such as social interaction[10,15] and interpersonal skills[10,11]. Evidence exists that regular dance activity could reduce the risk of falling[16,17] by improving sensorimotor function[13], balance, strength, mobility and flexibility[15,17-19]. A number of studies reported improvements in aerobic power, endurance, cardiovascular health[18,19] and metabolic disorders such as diabetes mellitus[20]. Indications are that bone-mineral content may benefit, but confirmation is needed.[19] As for other forms of exercise, it is feasible to surmise that the cellular and molecular anti-aging effects referred to in the introduction[4] also apply to dancing as physical exercise.

## Comparing the effects of dance to those of other forms of physical activity

Several studies found that dance, as a form of physical exercise, could be as effective as conventional structured exercise. However, variations in dance genre, dance protocols, aerobic intensity and methods of appraisal, confound a general comparison between the benefits of dance and those of other forms of physical activity. In a systematic review of seven control trials on the effects of dance on cardiovascular risks and aerobic capacity with aging, no significant differences were reported between the effects of dance and those of other types of exercise.[21] In a review of 28 studies, comprising 1276 participants, similar results were found for cardiovascular function and self-perceived mobility, while the effects of dance were occasionally reported as superior for certain musculoskeletal functions and blood parameters.[22] A pooled analysis of 11 longitudinal population-based British surveys on dance and cardiovascular disease mortality, showed that while both moderate-intensity dancing and moderate-intensity walking are inversely associated with cardiovascular disease mortality, greater risk reduction can be obtained with dancing.[23] Explanations for the reported superior benefits were that the social aspects of recreational dance could have acted as a buffer against psychosocial stress, and that moderate-intensity dancing often includes short bouts of vigorous intensity, thus mimicking high-intensity interval training.[23]

Results of studies that involve older participants in dance programmes marked by constant learning of new choreographies and movement sequences, suggest the long-term effects of such dance programmes are superior to the effects of either repetitive physical exercises or intensity-matched conventional fitness training in inducing cerebral neuroplasticity.[24] Dancing involves skills that are not necessarily required in other forms of exercise, such as coordination, learning of new movement sequences, integration of movement with music, as well as a number of expressive and communicative skills.[25] Dance thus integrates several different brain areas which then improves neuroplasticity.[26] Structural changes associated with dancing include increases in hippocampal volume, in grey matter volume of the left precentral and para-hippocampal gyrus, improved white matter integrity, as well as increased cortical thickness in the lateral occipito-temporal cortex, an area involved in action observation, visuomotor integration and action imitation – in other words, activities important for motor learning and executing skilled movements.[26,27]

Many organisations, including the World Health Organization (WHO), have recommended guidelines for physical activity for older people.[28] Most suggest 150–300 minutes of moderate-intensity exercise per week. This duration

most probably exceeds the time that older individuals spend on once-a-week dancing. However, while the more-is-better dose-response relationship is still maintained by many, recent indications are that significant health benefits can be gained by simply becoming physically more active.[29] The fact that dance is seen by many as highly enjoyable, and that enjoyment is known to be a strong predictor of perseverance with physical activity[24], probably contributes to a lower attrition rate.

## Potential risks

The main risks for older people participating in dance are probably vulnerability to falling, and effects on cardiovascular morbidity. Little information is available on the occurrence of serious falls in the elderly while participating in dance as a leisure activity. Nevertheless, as for other forms of physical exertion, guidance from medical practitioners could be valuable, especially for those with chronic conditions and/or functional limitations. Similarly to the risk of falling, hardly any records could be found on negative cardiovascular events. Nonetheless, it seems feasible to assume that general guidelines apply, whatever the exercise. According to the WHO, older adults with serious cardiac diseases or those recovering from an acute cardiac event, should rather be enrolled in a cardiac rehabilitation programme involving a multidisciplinary approach, with the prescribed physical activity component consisting of a medical assessment, cardiac risk factor management and psychosocial interventions.[30] Some evidence exists for dance, including low-impact dance, as a supplementary tool for cardiac rehabilitation.

## Spectrum of dance-based physical activities for older people

Dance for the older person varies from that on par with dance for the much younger, to creative dance classes in nursing homes and acute hospital settings. Intensity levels for group dance classes range from gentle movements in the seated position to dynamic dance. A variety of dance-based exercises exists, including dance movement therapy, dance therapy and programmes with physiotherapy falls-prevention activities integrated into dance programmes. For further information, see the supplementary material for examples of dance as exercise for older people.

## Conclusions

Dance, through its beneficial effects on physical and mental well-being, can potentially slow down the functional decline associated with aging. Many older people join dance groups for socialising and the joy of dancing, rather than the health benefits. Dancing, although not for everyone, could potentially contribute to meeting both the health and the social aspects. As with other forms of exercise, inappropriate dance intensity could be a risk factor for some, especially those with chronic conditions and/or functional limitations, in which case guidance from a medical practitioner is advisable.

## Competing interests

We have no competing interests to declare.

## References

1. Langhammer B, Bergland A, Rydwik E. The importance of physical activity exercise among older people. Biomed Res Int. 2018;2018, Art. #7856823. https://doi.org/10.1155/2018/7856823
2. Christmas C, Andersen RA. Exercise and older patients: Guidelines for the clinician. J Am Geriatr Soc. 2000;48(3):318–324. https://doi.org/10.1111/j.1532-5415.2000.tb02654.x
3. Papalia GF, Papalia R, Diaz Balzani LA, Torre G, Zampogna B, Vasta S, et al. The effects of physical exercise on balance and prevention of falls in older people: A systematic review and meta-analysis. J Clin Med. 2020;9(8):2595. https://doi.org/10.3390/jcm9082595
4. Rebelo-Marques A, De Sousa Lages A, Andrade R, Ribeiro CF, Mota-Pinto A, Carrilho F, et al. Aging hallmarks: The benefits of physical exercise. Front Endocrinol. 2018;9, Art. #258. https://doi.org/10.3389/fendo.2018.00258
5. López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G. The hallmarks of aging. Cell. 2013;153(6):1194–1217. https://doi.org/10.1016/j.cell.2013.05.039
6. Hewston P, Kennedy CC, Borhan S, Merom D, Santaguida P, Ioannidis G, et al. Effects of dance on cognitive function in older adults: A systematic review and meta-analysis. Age Ageing. 2021;50(4):1084–1092. https://doi.org/10.1093/ageing/afaa270
7. Meng X, Li G, Jia Y, Liu Y, Shang B, Liu P, et al. Effects of dance intervention on global cognition, executive function and memory of older adults: A meta-analysis and systematic review. Aging Clin Exp Res. 2020;32(1):7–19. https://doi.org/10.1007/s40520-019-01159-w
8. Zhu Y, Zhong Q, Ji J, Ma J, Wu H, Gao Y, et al. Effects of aerobic dance on cognition in older adults with mild cognitive impairment: A systematic review and meta-analysis. J Alzheimers Dis. 2020;74(2):679–690. https://doi.org/10.3233/JAD-190681
9. Koch SC, Riege RFF, Tisborn K, Biondo J, Martin L, Beelmann A. Effects of dance movement therapy and dance on health-related psychological outcomes. A meta-analysis update. Front Psychol. 2019;10:1806. https://doi.org/10.3389/fpsyg.2019.01806
10. Jiménez J, Bräuninger I, Meekums B. Dance movement therapy with older people with a psychiatric condition: A systematic review. Arts Psychother. 2019;63:118–127. https://doi.org/10.1016/j.aip.2018.11.008
11. Predovan D, Julien A, Esmail A, Bherer L. Effects of dancing on cognition in healthy older adults: A systematic review. J Cogn Enhanc. 2019;3:161–167. https://doi.org/10.1007/s41465-018-0103-2
12. McNeely ME, Duncan RP, Earhart GM. Impacts of dance on non-motor symptoms, participation, and quality of life in Parkinson disease and healthy older adults. Maturitas. 2015;82(4):336–341. https://doi.org/10.1016/j.maturitas.2015.08.002
13. Kshtriya S, Barnstaple R, Rabinovich DB, De Souza JFX. Dance and aging: A critical review of findings in neuroscience. Am J Dance Ther. 2015;37:81–112. https://doi.org/10.1007/s10465-015-9196-7
14. Marks R. Narrative review of dance-based exercise and its specific impact on depressive symptoms in older adults. AIMS Med Sci. 2015;3(1):61–76. https://doi.org/10.3934/medsci.2016.1.61
15. Guzmán-García A, Hughes JC, James IA, Rochester L. Dancing as a psychosocial intervention in care homes: A systematic review of the literature. Int J Geriatr Psychiatr. 2013;28(9):914–924. https://doi.org/10.1002/gps.3913
16. Mattle M, Chocano-Bedoya PO, Fischbacher M, Meyer U, Abderhalden LA, Lang W, et al. Association of dance-based mind-motor activities with falls and physical function among healthy older adults: A systematic review and meta-analysis. JAMA Netw Open. 2020;3(9), e2017688. https://doi.org/10.1001/jamanetworkopen.2020.17688
17. Fernández-Argüelles EL, Rodríguez-Mansilla J, Antunez LE, Garrido-Ardila EM, Muñoz RP. Effects of dancing on the risk of falling related factors of healthy older adults: A systematic review. Arch Gerontol Geriatr. 2015;60(1):1–8. https://doi.org/10.1016/j.archger.2014.10.003
18. Liu X, Shen PL, Tsai YS. Dance intervention effects on physical function in healthy older adults: A systematic review and meta-analysis. Aging Clin Exp Res. 2021;33(2):253–263. https://doi.org/10.1007/s40520-019-01440-y
19. Keogh JW, Kilding A, Pidgeon P, Ashley L, Gillis D. Physical benefits of dancing for healthy older adults: A review. J Aging Phys Act. 2009;17(4):479–500. https://doi.org/10.1123/japa.17.4.479
20. Rodrigues-Krause J, Krause M, Reischak-Oliveira A. Dancing for healthy aging: Functional and metabolic perspectives. Altern Ther Health Med. 2019;25(1):44–63.
21. Rodrigues-Krause J, Farinha JB, Krause M, Reischak-Oliveira Á. Effects of dance interventions on cardiovascular risk with ageing: Systematic review and meta-analysis. Complement Ther Med. 2016;29;16–28. https://doi.org/10.1016/j.ctim.2016.09.004
22. Fong Yan A, Cobley S, Chan C, Pappas E, Nicholson LL, Ward RE, et al. The effectiveness of dance interventions on physical health outcomes compared to other forms of physical activity: A systematic review and meta-analysis. Sports Med. 2018;48(4):933–951. https://doi.org/10.1007/s40279-017-0853-5

23. Merom D, Ding D, Stamatakis E. Dancing participation and cardiovascular disease mortality: A pooled analysis of 11 population-based British cohorts. Am J Prev Med. 2016;50(6):756–760. https://doi.org/10.1016/j.amepre.2016.01.004

24. Rehfeld K, Lüders A, Hökelmann A, Lessmann V, Kaufmann J, Brigadski T, et al. Dance training is superior to repetitive physical exercise in inducing brain plasticity in the elderly. PLoS ONE. 2018;13(7), e0196636. https://doi.org/10.1371/journal.pone.0196636

25. Predovan D, Julien A, Esmail A, Bherer L. Effects of dancing on cognition in healthy older adults: A systematic review. J Cogn Enhanc. 2019;3:161–167. https://doi.org/10.1007/s41465-018-0103-2

26. Teixeira-Machado L, Arida RM, de Jesus Mari J. Dance for neuroplasticity: A descriptive systematic review. Neurosci Biobehav Rev. 2019;96:232–240. https://doi.org/10.1016/j.neubiorev.2018.12.010

27. Rektorova I, Klobusiakova P, Balazova Z, Kropacova S, Sejnoha Minsterova A, Grmela R, et al. Brain structure changes in nondemented seniors after six-month dance-exercise intervention. Acta Neurol Scand. 2020;141(1):90–97. https://doi.org/10.1111/ane.13181

28. The World Health Organization (WHO). WHO guidelines on physical activity and sedentary behaviour. Geneva: WHO; 2020.

29. Warburton DER, Bredin SSD. Health benefits of physical activity: A systematic review of current systematic reviews. Curr Opin Cardiol. 2017;32(5):541–556. https://doi.org/10.1097/HCO.0000000000000437

30. Liew JM, Teo SP. Physical activity in older people with cardiac co-morbidities. J Geriatr Cardiol. 2018;15(8):557–558. https://doi.org/10.11909/j.issn.1671-5411.2018.08.004

**AUTHOR:**
Jennifer M. Fitchett[1]

**AFFILIATION:**
[1]School of Geography, Archaeology and Environmental Studies, University of the Witwatersrand, Johannesburg, South Africa

**CORRESPONDENCE TO:**
Jennifer Fitchett

**EMAIL:**
jennifer.fitchett@wits.ac.za

**HOW TO CITE:**
Fitchett JM. Science always makes a difference. S Afr J Sci. 2021;117(11/12), Art. #12471. https://doi.org/10.17159/sajs.2021/12471

**ARTICLE INCLUDES:**
☐ Peer review
☐ Supplementary material

**KEYWORDS:**
knowledge creation, science as a public good, science engagement

**PUBLISHED:**
29 November 2021

# Science always makes a difference

How do we know if science makes a difference? This question was posed to participants of a webinar hosted by the *South African Journal of Science* during the 2021 National Science Week. A report of the webinar has been provided by Masete[1]. As a panellist in this webinar, I argued that science always makes a difference; and that therefore there is no need for an approach to quantify or detect or measure this difference. Indeed, attempting to separate out the science that makes a difference from that which is perceived to not make a difference does a disservice to the advancement of science. Science always makes a difference, because intrinsic to science is the pursuit of new knowledge. Even if this new knowledge is small in magnitude or influence, parochial, or restricted to a narrow subdiscipline, even if it is seldom cited, it provides one of many critical building blocks for future work. Here I summarise the position I outlined in the webinar, through arguing (1) the cases in which the 'difference' made by science is most tangible, (2) the more continuous impacts of scientific pursuit that likewise make a difference, and (3) that the difference that science makes does not always contribute a net good to society.

The 'difference' made by science is most tangible when it does lead to a breakthrough, a discovery or a solution to a problem. The difference made by science is clear and detectable when it involves raising awareness about a problem, improving the understanding of an issue, or improving the accuracy of models or forecasts. The difference that science makes is indisputable when it is seen to encourage young people to pursue further education, particularly in STEM fields, or when it addresses social ills. These differences are often the types of tangible outcomes that students pursuing STEM degrees expect to be able to contribute towards after graduation, and certainly are tremendously valuable.

However, I would argue that science equally makes a difference on a far more continuous basis, and at a far smaller scale. Conducting scientific research – following, engaging with, developing, refining, and repeating the scientific method makes a difference. Training young scientists, whether in the classroom or lecture theatre, the laboratory or the field, and whether in lectures or through demonstration, makes a difference. Communicating research findings, whether to an academic or public audience, whether in writing or verbally, whether to a small or large group, makes a difference. Using research to inform policy, without doubt, makes a difference.

The 'difference' that science makes, however, is not always for good, or for good in perpetuity. Science can also drive some of the world's greatest problems. What might be developed as a solution to a problem now, might become detrimental down the line. Mistakes are made in science, with catastrophic consequences. We also cannot ignore the role that science has played in some of the world's worst wars. In promoting science for good, we need to regularly reflect on the importance of research integrity, of ethical practices, of rigour in our approach, and in transparency in our procedures.

So if the question of 'when science makes a difference' is so easily answered by 'always', as I argue it is, what other questions should scientists be asking? I would posit:

1. What kind of difference do I, as a scientist, want to make?
2. How will I measure, quantify or detect the difference that I am making?
3. How will I determine whether the difference I have made provides a net positive outcome (however small or big) for the world now, and whether it will likely continue to do more good than harm in future?
4. How can I improve the impact of my science?

As a public, whether involved in the sciences or not, and whether working with a specific scientific subdiscipline or not, we also play an important role in recognizing the difference that science can make. We too need to self-reflect, and I argue that the following questions are important considerations:

1. How do we support science through recognising, and celebrating, the less tangible impacts?
2. How can we encourage 'science for good', or science which has a net positive impact?
3. How do we facilitate more people working in the sciences seeing and appreciating their value, and the difference that they make?
4. How can we best understand and learn from the outputs of science?

In recognising that science always makes a difference, the importance of science for good and science with good intention is heightened.

## Reference

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

**AUTHORS:**
Francois A. Engelbrecht[1]
Pedro M.S. Monteiro[2]

**AFFILIATIONS:**
[1]Global Change Institute, University of the Witwatersrand, Johannesburg, South Africa
[2]Southern Ocean Carbon – Climate Observatory (SOCCO), Council for Scientific and Industrial Research (CSIR), Cape Town, South Africa

**CORRESPONDENCE TO:**
Francois Engelbrecht

**EMAIL:**
Francois.Engelbrecht@wits.ac.za

**HOW TO CITE:**
Engelbrecht FA, Monteiro PMS. The IPCC Assessment Report Six Working Group 1 report and southern Africa: Reasons to take action. S Afr J Sci. 2021;117(11/12), Art. #12679. https://doi.org/10.17159/sajs.2021/12679

**ARTICLE INCLUDES:**
☐ Peer review
☐ Supplementary material

**KEYWORDS:**
climate change, coupled carbon-climate systems, net zero $CO_2$ emissions, southern Africa

**PUBLISHED:**
29 November 2021

# The IPCC Assessment Report Six Working Group 1 report and southern Africa: Reasons to take action

The release of the Intergovernmental Panel on Climate Change (IPCC) Assessment Report Six (AR6) Working Group I (WG1) report in August 2021 brought to completion what is arguably the most thorough and scrutinised assessment of climate knowledge needed to steer the planet away from dangerous and irreversible climate change.[1] The AR6 WG1 report builds on the Assessment Report Five WG1 report published in 2013, as well as on three special reports commissioned during the AR6 cycle: the *Special Report on Global Warming of 1.5 °C* (SR1.5), the *Special Report on the Ocean and Cryosphere in a Changing Climate*, and the *Special Report on Climate Change and Land*. The AR6 WG1 report as such does not come up with fundamentally new insights into the planet's coupled carbon-climate systems, but through its assessment of ~14 000 publications and response to 78 000 review comments, it provides an unprecedented level of confidence to earlier findings. What it reveals is that there have been critical methodological advances in both observation and modelling that have enabled improved levels of confidence[2] and opened doors for new science on global and regional climate-carbon challenges. These improved confidence levels are necessary to support the deeply transformative global, and South African, decision-making towards net zero $CO_2$ emissions by 2050 in support of restricting global warming to below 1.5 °C relative to the pre-industrial temperatures.[3,4] Our Commentary focuses on the assessment of the global climate-carbon system with implications for adaptation and mitigation action in southern Africa, and we provide a view of new research opportunities for regional climate and sustainability science.

## Dangerous and irreversible climate change may already be unavoidable

Five illustrative scenarios of future greenhouse gas emissions underpin the AR6 WG1 report. These are referred to as Shared Socio-economic Pathways (SSPs) and range from high (SSP1-1.9 and SSP2-2.7) to low (SSP3-7.0 and SSP5-8.5) mitigation futures.[1] SSP1-1.9 may be described as a 'best effort' mitigation scenario. In this scenario, $CO_2$ emissions are cut by about half by the year 2030, compared to present-day levels. Thereafter $CO_2$ emission reductions continue to be implemented, with net-zero emissions (that is, further increases in atmospheric $CO_2$ concentrations cease) reached by 2050 within a required remaining carbon budget.[1,5]

The report assesses that the best estimate of global warming (the increase in the average surface temperature of the planet) for the period 2011–2020 with respect to pre-industrial temperature is 1.1 °C[1], approaching the thresholds of 1.5 °C and 2 °C that define 'dangerous climate change' in the Paris Agreement on Climate Change. At these levels of global warming, aspects of climate change such as drastic sea-level rise may become irreversible, whilst extreme weather events may occur at unprecedented frequencies and intensities. A startling assessment of the report is that it is more likely than not (the probability is greater than 50%) that the 1.5 °C threshold of global warming will be exceeded by a small margin even with the best-effort mitigation, and that this exceedance may occur as soon as the early 2030s.[1,6] Exceedance is defined in terms of the mid-point of a 20-year long period of average global surface temperature[6], implying that the world may already be in the first 20-year period for which the average surface temperature will be 1.5 °C higher than the pre-industrial baseline. This is perhaps one reason why the AR6 WG1 report has been described as a 'code red' for humanity by the United Nations Secretary General – the notion that exceeding at least the lower threshold defining dangerous climate change may no longer be avoidable. Humanity seems to have delayed strong climate action for too long, and the window to avoid dangerous or even catastrophic impacts may already be closed.

The report, however, also assesses that, with best-effort mitigation specifically under SSP1-1.9, the chance of restricting global warming to below 2 °C is excellent.[1,6] Moreover, the chances are good that, even if an overshoot of the 1.5 °C level occurs, global warming can be reduced again in the second half of the century to below this threshold, in the presence of $CO_2$ removal technologies (still to be developed at scale). Certainly, at least to some extent, the level of future global warming and the range of future climate change impacts remain in the hands of humanity.

The AR6 goes to some lengths to reaffirm why the 1.5 °C and 2 °C levels of global warming are thought to be critical to avoid. These levels of global warming are thresholds at which global tipping points may be reached and aspects of climate change may become irreversible.[7] First and foremost is the potential instability of both the Greenland and West Antarctic ice sheets – the assessment being that irreversible collapse of both sheets may be triggered by sustained global warming of 1.5–2 °C.[7] Complete loss of the ice sheets will likely take place at timescales of millennia. Sustained global warming of 2 °C is assessed to be associated with 6 m of sea-level rise over a period of 2000 years – implying that a lack of climate action today will commit future generations to a completely different coastline and reduced living space. Moreover, the report assesses, that under low mitigation futures, sea-level rise may reach levels of about 1–1.9 m by 2150[1,7] – sufficient to displace hundreds of millions of people from coastal areas[8]. It should be realised that sea-level rise is fundamentally irreversible across the lifetimes of many generations.

## Every bit of global warming matters

Increasingly, climate change attribution science is capable of quantifying the role of human influence in the occurrence of individual weather events (as opposed to these events occurring entirely because of natural

variability). For example, the June 2021 heatwave in the Pacific Northwest has been assessed to have been impossible without the effects of anthropogenic warming[9], whilst climate change made the flooding that occurred in Germany in July 2021 up to nine times more likely to occur[10]. The AR6 WG1 report points out that increases in extreme weather events can already be detected across all regions of the world, and that these increases can be attributed to human-induced global warming.[1] Global warming of 1.5–2 °C brings a commitment to increases in the frequency and intensity of a multitude of extreme weather events, such as heatwaves, heavy precipitation, intense tropical cyclones and droughts in some regions.[11] Such changes are referred to as 'dangerous climate change'. The report is clear that increases in extreme weather events, both in terms of frequency and intensity, are higher at 2 °C than at 1.5 °C global warming, with further increases as the level of global warming increases.[11] That is, even if a best-effort mitigation fails to restrict global warming to below 1.5 °C, there will still be substantial benefits in limiting the overshoot to under 2 °C, and then bringing the temperature down through negative emissions.[5,6,12]

## Southern Africa as a climate change hotspot and the potential for regional tipping points

The AR6 WG1 report confirms previous IPCC assessments: the southern African region is likely to become drier, even under 1.5 °C of global warming (Figure 1).[6] Moreover, the observed rate of warming in recent decades is about twice the global average[13], and further drastic warming is projected for the region as the level of global warming increases (Figure 2)[6]. This renders the region a climate change hotspot[14]: when a region that is naturally dry and warm becomes drastically warmer and drier, the options for adaptation are limited.

The report assesses that it is not only general reductions in precipitation that are likely over southern Africa in a warmer world, droughts are also projected to occur more frequently, even under 1.5 °C of global warming. Moreover, as the level of global warming increases, so does the frequency and intensity of drought.[11] Over the eastern part of southern Africa, increases in heavy precipitation are projected in a warmer world, despite the region projected to become generally drier.[11] Against this general assessment from the report, it is important to consider the possibility of the existence of regional tipping points in the southern African climate system, where a new climate regime is reached and high-impact climate events unprecedented in this historical record start to occur.

The single biggest climate change risk that South Africa may have to face in the near term (2021–2040), is one or more 'day zero' droughts occurring in the Gauteng Province. Such an event, where the portion of Gauteng's water supply from the eastern mega dams is severely compromised by a multi-year drought, may be devastating to the South African economy. Gauteng is the economic heartland of the country where most of the industries are located, and where 16 million people live. It may be noted that in September 2016, the level of the Vaal Dam in South Africa fell to about 26%[15] at the end of a 4-year drought. The drought culminated with the 2015/2016 El Niño, one of the strongest ever recorded[16], and brought the realisation that the risk of a 'day zero' drought in Gauteng already exists. That is, the level of the dam came close to falling to below 20% – the threshold at which the Gauteng water supply would have been severely compromised. Multi-year droughts have also impacted severely on South Africa's southern coastal cities Gqeberha and Cape Town in recent years[17], and, in the latter case, it has been assessed that climate change has already increased the likelihood of 'day zero' type droughts to occur by a factor of three[18].

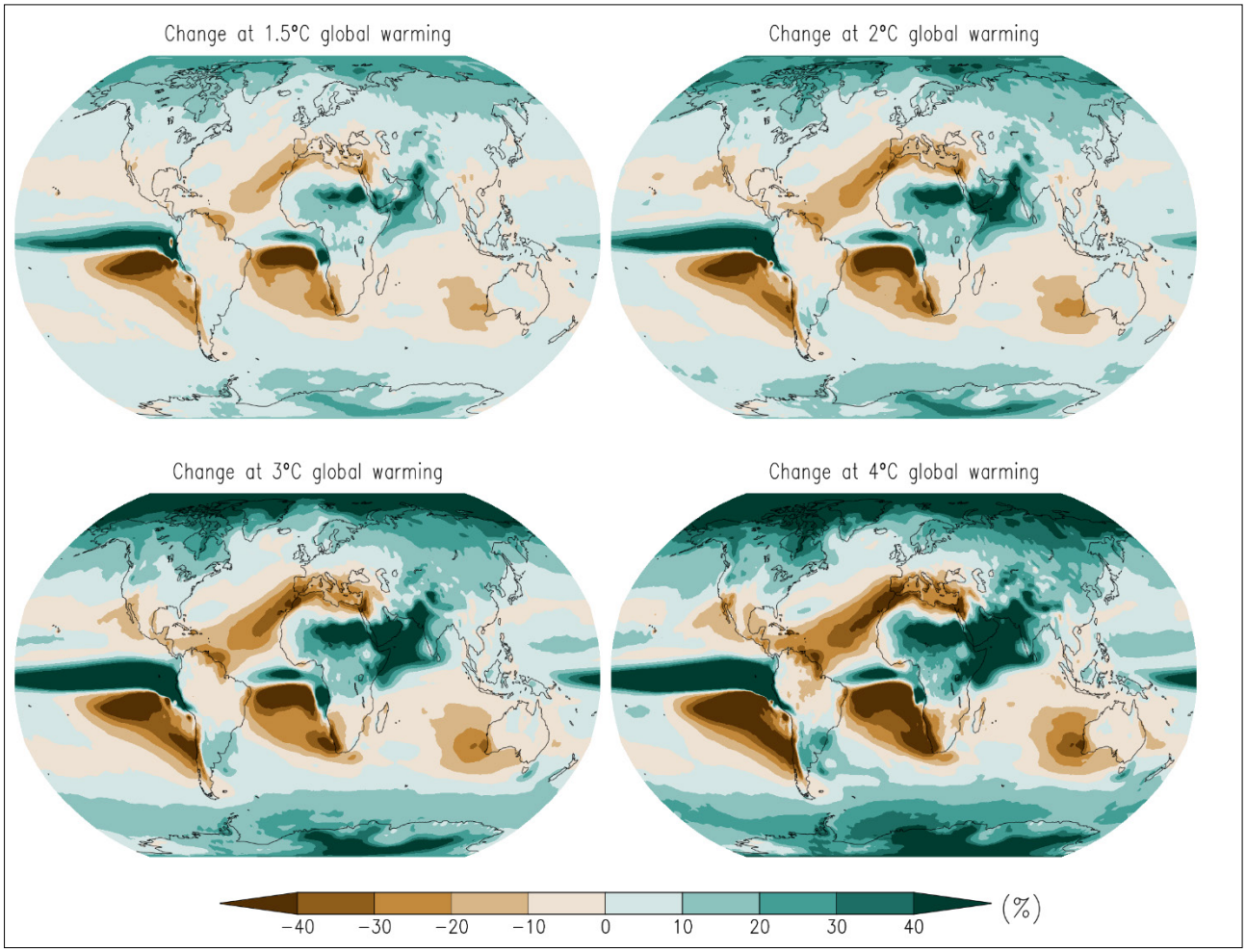



**Figure 1:** Projected spatial patterns of change in annual average precipitation (expressed as a percentage change) at different levels of global warming (1.5 °C, 2 °C, 3 °C, and 4 °C) relative to the period 1850–1900. Values were assessed from a 20-year period at a given warming level, based on model simulations under SSP5-8.5 by 30 global circulation models that contributed to the Coupled Model Intercomparison Project Phase Six (CMIP6).

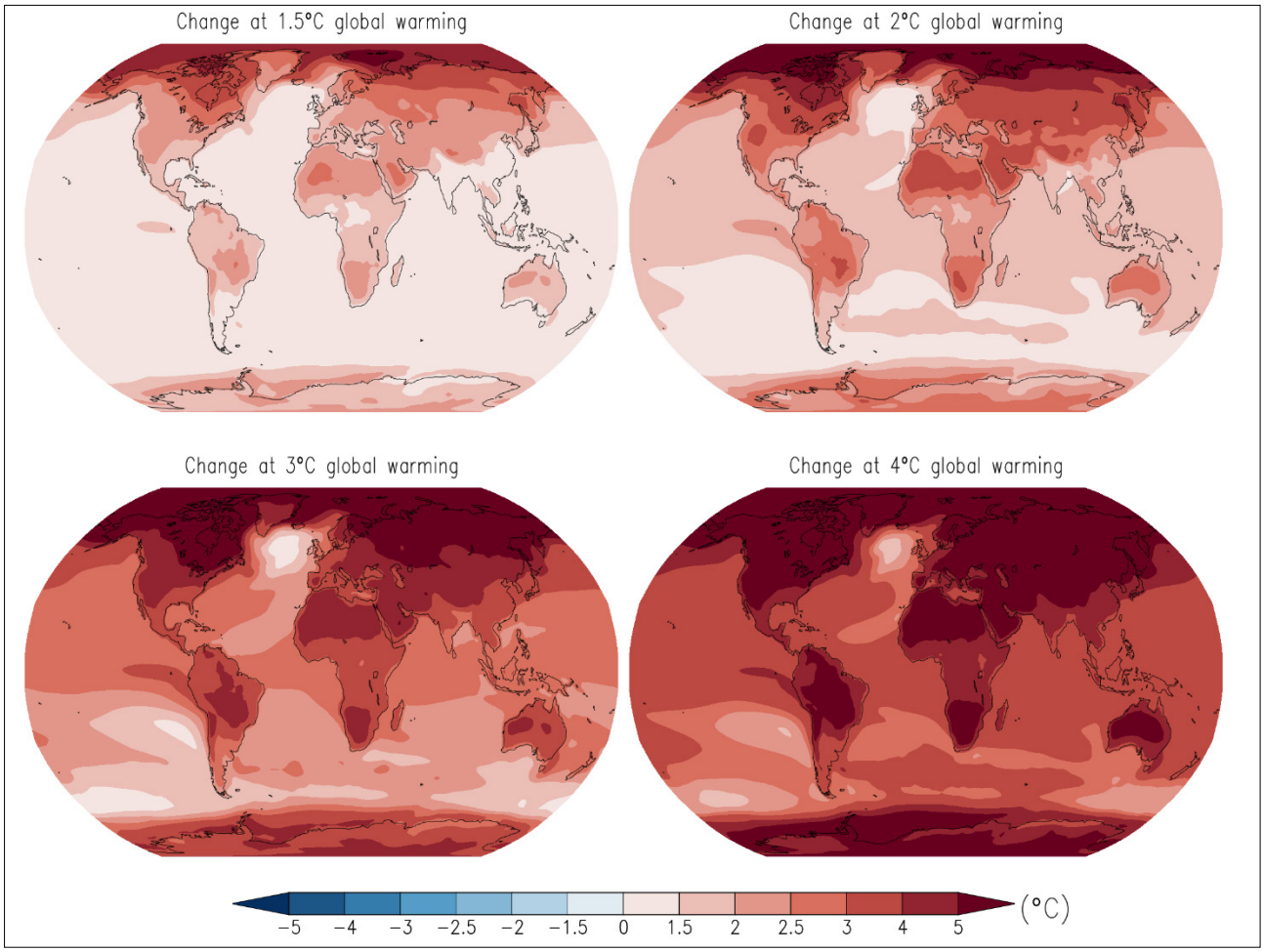



**Figure 2:** Projected spatial patterns of change in surface temperature (°C) at different levels of global warming (1.5 °C, 2 °C, 3 °C, and 4 °C) relative to the period 1850–1900. Values were assessed from a 20-year period at a given warming level, based on model simulations under SSP5-8.5 by 30 global circulation models that contributed to the Coupled Model Intercomparison Project Phase Six (CMIP6).

The possibility therefore exists that South Africa's climate may drift into an entirely new regime, where 'day zero' droughts become regular events, substantially impacting on the sustainable growth of the biggest cities. Generating probabilistic projections to better quantify the magnitude of this risk needs to be a priority area of research for South African climatologists and hydrologists.

Periods of drought are also associated with increasing heatwave frequencies in southern Africa. Heatwaves are certain to occur at unprecedented intensities in a warmer world, with direct implications for unprecedented impacts on human mortality in many parts of the world, including southern Africa.[14,19] The combination of more frequent heatwaves and multi-year droughts may be devastating to agriculture, local livelihoods and food production in the region. The IPCC SR1.5 assessed that, under 3 °C of global warming, both the maize crop and the cattle industry in southern Africa are likely to collapse, largely because of the biophysical impacts of increased heat stress in a drier climate.[14] What is less well understood, is whether this tipping point may be reached at even lower levels of global warming, due to the economic impacts of drought on farmers, at least in those areas in the region where farming is already marginal. Finally, in Mozambique and northeastern South Africa, the potential for a completely different type of tipping points exists. In a warmer world, intense tropical cyclones are likely to occur more frequently[11], and the possibility exists that these systems can follow more poleward trajectories[20]. It is possible that an intense tropical cyclone (equivalent to a category three to five hurricane) can make landfall at Maputo, or even further to the south, at Richards Bay. More research is needed to better quantify how the risk of such an event occurring depends on the level of global warming.

A special research focus is needed to develop a probabilistic perspective on the occurrence of regional tipping points in the southern African climate-ecosystem-human-system nexus. This should include both a biophysical perspective as well as take into account how socio-economics may accelerate the potential collapse of existing industries and systems. Even if these events are found to be of low probability, they are potentially of such high – or even devastating impact – that they need to be considered in risk assessments for adaptation planning and, in particular, disaster risk reduction strategies.

Improving the confidence in climate change projections depends strongly on a good understanding of the global coupled carbon-climate system and its feedbacks under increasing, decreasing and negative emissions.[1] Carbon-climate linkages are discussed in more detail in the following sections, again through a southern African lens.

## The past, present and future of carbon sinks and emerging feedbacks

Carbon is the main lever by which to strengthen or weaken global warming and climate change – it sets the rate of global and regional warming through a logarithmic relationship between concentration of $CO_2$ in the atmosphere and resulting radiative forcing.[12] The AR6 WG1 report assessed that it is unequivocal that increasing greenhouse gases in the atmosphere are linked to human activities.[1,5] However, the rate at which $CO_2$ builds up in the atmosphere depends both on emissions as well as on the magnitudes of the ocean and land carbon sinks.[5] Understanding and predicting the trends and variability of the land and ocean sinks and their sensitivity to both carbon and climate feedbacks is

therefore essential to improving the confidence of future climate change projections under low and high mitigation scenarios.[5,6,21]

In the past six decades, the large negative feedbacks of the ocean and terrestrial carbon sinks have slowed global warming by partitioning anthropogenic $CO_2$ emissions by ~23% and ~27%, respectively, leaving a quasi-steady residual of 44% in the atmosphere. It is this airborne fraction that drives the increasing $CO_2$ concentrations and global warming.[5] Palaeo and historical observations, together with significantly improved models, have provided significantly greater confidence in the assessment of the constraints of the global carbon budget and the sensitivity of the sinks to both carbon and climate feedbacks. Assessed paleo observations highlighted that although the magnitudes of $CO_2$ in the atmosphere are unprecedented in the past 1 Myr, they scale to the magnitudes preceding the late Miocene period, 3–5 Myr ago.[5,21] What is unprecedented about the historical $CO_2$ emissions are the rates at which contemporary $CO_2$ is increasing in the atmosphere, which have been assessed to be at least an order of magnitude higher than at any time in the past 60 Myr.[5] The contemporary global carbon-climate system is thus in a far more perturbed transient state than was the case in the pre-industrial and most of the palaeo periods. Notwithstanding the exponential rate of increase in historical global $CO_2$ emissions, the land and ocean carbon sinks slowed global warming by taking up about a combined 50% of $CO_2$ emissions over the past 60 years.[5] These negative feedbacks are mainly driven by the physics and $CO_2$ chemistry in the ocean and the biosphere on land.[5] Observations show that carbon and climate positive feedbacks that could increase the $CO_2$ fraction in the atmosphere are still too small. However, observations also show that processes that may influence future positive feedbacks are undergoing observable changes in both the ocean and on land. Decreasing ocean buffering capacity, ocean acidification, and ocean warming as well as warming and rainfall extremes on land serve as examples.[5]

Model projections show that ocean and land carbon sinks will weaken across all scenarios in the second half of the century, but the processes that drive this weakening are quite different between high and low mitigation pathways.[21] Whereas under high mitigation the weakening ocean and land sinks are linked to a decreasing atmospheric $CO_2$ arising from negative emissions, under low mitigation scenarios the weakening is driven by the growing positive carbon-concentration and carbon-climate positive feedbacks on both land and the ocean.[21] One of the outcomes of these different sink responses is that the stronger the emission scenario, the smaller the fraction of emitted $CO_2$ that is taken up by the sinks. This increases the rate at which emitted $CO_2$ builds up in the atmosphere as the airborne fraction. The full understanding of the mechanistic basis of the feedback responses in the ocean and on land under both high and low emissions remains an important research challenge to improve the confidence in model projections.

On the global spatial scale, projections provide an indication of the geographical heterogeneity of the sensitivity of ocean and land carbon sinks to climate and carbon forcing (Figure 3). These show that the negative feedback of land and ocean carbon sinks to rising atmospheric $CO_2$ are offset by positive feedback responses to climate forcing (warming and rainfall) and decreasing buffering capacity in the ocean. Importantly it points to the global role of southern hemisphere tropical and sub-tropical terrestrial ecosystems as well as the Southern Ocean, both of which highlight southern Africa's underutilised comparative geographic advantage for regional and global climate and earth system science research and climate risk projection.[5,21]

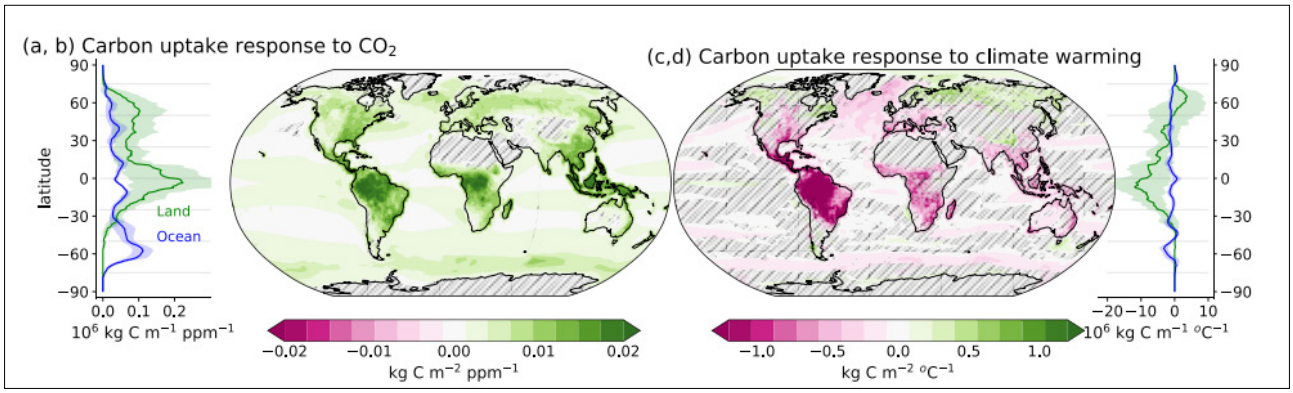
## Slowing down and stopping warming – every ton of $CO_2$ adds to global warming

One of the most important metrics, which sets the scientific basis for the need to achieve net-zero $CO_2$ emissions within a specific remaining carbon emissions budget to remain below a specific warming level by 2050, is the Transient Climate Response to cumulative carbon Emissions (TCRE).[1,3,5] It is a quasi-linear relationship between warming and cumulative $CO_2$ emissions, which shows that to arrest additional warming, the emissions-driven increase in atmospheric $CO_2$ must stop completely. It also provides the science basis to constrain the remaining carbon budget that can be emitted for a given probability to remain below a chosen global mean surface temperature target; for example, 1.5 °C or 2 °C.[4,5]

This transient relationship between warming and cumulative emissions is a property of the planet's carbon-climate system. It is largely regulated by the ocean's capacity to take up heat and the unique chemical properties of $CO_2$ that enables it to be stored in the ocean in much greater quantities as a salt rather than as a dissolved gas.[5] Again, the role of the Southern Ocean in sustaining this relationship under positive and negative emissions continues to present a global science challenge, which also highlights South Africa's comparative geographical advantage in respect of Antarctica and the Southern Ocean carbon-climate science (Figure 4).

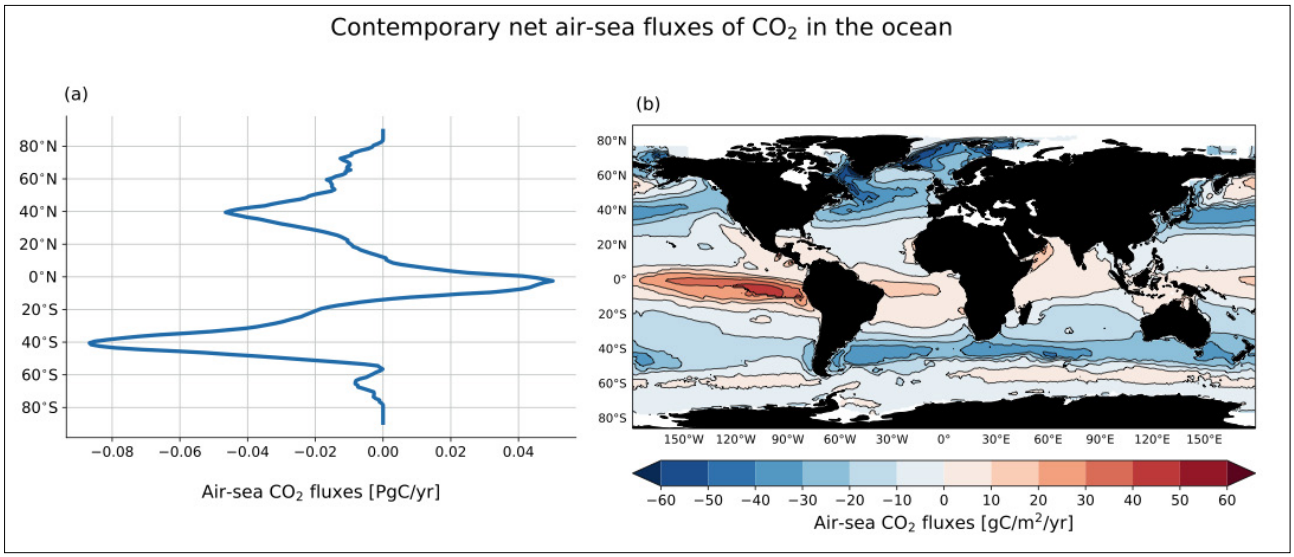
## Negative emissions – carbon dioxide removal

To restrict global warming to below the threshold of 1.5 °C by the end of the century, especially in the presence of an overshoot in the surface air temperature above 1.5 °C in the coming decades, lies the larger global and regional science and technology challenge of negative emissions. This removal of $CO_2$ from the atmosphere is necessary to offset the warming, the Zero Emissions Commitment (ZEC), associated with atmospheric $CO_2$ from historical emissions and to manage possible 1.5 °C temperature overshoot scenarios.[5,23,24]




*Source: From Box TS.5 Figure 1 of Arias et al.[21]*

**Figure 3:** Global and regional carbon cycle responses to carbon and climate forcing. Projected spatial and zonally integrated responses of terrestrial and ocean carbon sinks to carbon (a,b) and climate (c,d) forcing under a scenario of 1% increase in atmospheric $CO_2$ per year. It highlights that the ocean responses are mainly in the southern hemisphere mid- to high latitudes (blue in a,d) and terrestrial are mainly in the tropics and subtropics (green in a,d).




*Source: Djeutchouang L, personal communication using the South African CSIR-ML6 machine learning model for global ocean $CO_2$ used in AR6 WG1*[5,22]

**Figure 4:** Contemporary mean state of the global net ocean $CO_2$ fluxes (2008–2018) (sinks and sources) and showing the prominent role of the Southern Ocean (south of 30°S) as a global ocean $CO_2$ sink.


*Photo: Kristi Ebi*

**Figure 5:** The IPCC *Special Report on Global Warming of 1.5 °C* was released under the banner of 'Reasons for Hope' in October 2020, indicating that strong climate action could still restrict global warming to below 1.5 °C. In 2021, that hope has faded.

The impacts and effectiveness of $CO_2$ removal present enormous ethical, governance and technological scaling-up challenges.[5] However, they will be critical to avoiding catastrophic costs of climate extremes. The opportunity for South African science and technology innovation lies in the reality that, at this stage, there are no proven methods which are both technologically feasible and scalable to have the regional or global impact on atmospheric $CO_2$ needed to arrest warming and ameliorate ocean acidification. This is an enormous challenge and opportunity for large interdisciplinary consortia of South African science and technology innovation systems to address with directed public and private funding.

## Reasons for despair, hope, or action?

When the IPCC released the SR1.5 report in 2018, it framed the report under the message of 'Reasons for Hope' (Figure 5), pointing out that strong climate action could still prevent the dangerous threshold of global warming of 1.5 °C being reached. The SR1.5 report proceeded to outline the substantial benefits that exist in restricting global warming to 1.5 °C, as opposed to 2 °C, in terms of reduced impacts. In this respect, the assessment of the AR6 WG1 report is sobering: it is now considered more likely than not that the 1.5 °C threshold will be exceeded (as soon as the early 2030s, and probably by a small margin under best-effort mitigation). The realisation that the world is committed to further warming and further increases in extreme events, and possibly even elements of 'dangerous climate change' should, however, not lead to despair. Rather, the AR6 WG1 report should be seen as a call for strong climate change action, in terms of both mitigation and adaptation.

Best-effort mitigation requires strong, rapid and sustained cuts in greenhouse gas emissions towards net-zero global emissions within the prescribed remaining carbon budget, followed by negative emissions in the second half of the century. The pace and magnitude at which $CO_2$ emissions will have to be reduced for global warming to be restricted to 1.5 °C is staggering, and will require the formation of the strongest climate pact to date at the 26th Conference of the Parties (COP26) in Glasgow in November 2021. It will also require a concerted national and local mobilisation and sets of institutional architectures and actions to implement change. South Africa will be faced with tremendous challenges towards contributing fairly to global mitigation, whilst taking advantage of new investment opportunities to help ensure that its own transition away from fossil fuels is a just transition. The risks posed by future regional climate change impacts need to be quantified to the best level of confidence to formulate actionable messages for adaptation, and to pursue those actions with the same vigour as climate change mitigation. Of particular importance in this regard, is to develop improved probabilistic understanding of the potential for regional tipping points being reached in the southern African climate-ecosystem-human-system nexus, in the near term and beyond. The benefits of avoiding social and economic damage through immediate and effective mitigation, will minimise the costs of adaptation and support a just transition towards a low-carbon modern economy. South African science and technological innovation systems should be key partners in these challenges.

## Acknowledgements

We thank two anonymous readers for their comments, which helped to improve the clarity of the discussion.

## Competing interests

We have no competing interests to declare. F.A.E. and P.M.S.M. are Members of Working Group I of Assessment Report Six.

## References

1. IPCC. Summary for policymakers. In: Masson-Delmotte V, Zhai P, Pirani A, Connors SL, Péan C, Berger S, et al., editors. Climate change 2021: The physical science basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge, UK: Cambridge University Press. Forthcoming 2021.
2. Chen D, Rojas M, Samset BH, Cobb K, Diongue Niang A, Edwards P, et al. Framing, context, and methods. In: Masson-Delmotte V, Zhai P, Pirani A, Connors SL, Péan C, Berger S, et al., editors. Climate change 2021: The physical science basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge, UK: Cambridge University Press. Forthcoming 2021.
3. Rogelj J, Popp A, Calvin KV, Luderer G, Emmerling J, Gernaat D, et al. Scenarios towards limiting global mean temperature increase below 1.5 °C. Nat Clim Change. 2018;8:325–332. https://doi.org/10.1038/s41558-018-0091-3
4. Warszawski L. All options, not silver bullets, needed to limit global warming to 1.5°C: A scenario appraisal. Environ Res Lett. 2021;16(6), Art. #064037. https://doi.org/10.1088/1748-9326/abfeec
5. Canadell JG, Monteiro PMS, Costa MH, Cotrim da Cunha L, Cox PM, Eliseev AV, et al. Global carbon and other biogeochemical cycles and feedbacks. In: Masson-Delmotte V, Zhai P, Pirani A, Connors SL, Péan C, Berger S, et al., editors. Climate change 2021: The physical science basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge, UK: Cambridge University Press. Forthcoming 2021.
6. Lee JY, Marotzke J, Bala G, Cao L, Corti S, Dunne JP, et al. Future global climate: scenario-based projections and near-term information. In: Masson-Delmotte V, Zhai P, Pirani A, Connors SL, Péan C, Berger S, et al., editors. Climate change 2021: The physical science basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge, UK: Cambridge University Press. Forthcoming 2021.
7. Fox-Kemper B, Hewitt HT, Xiao C, Aðalgeirsdóttir G, Drijfhout SS, Edwards TL, et al. Ocean, cryosphere and sea level change. In: Masson-Delmotte V, Zhai P, Pirani A, Connors SL, Péan C, Berger S, et al., editors. Climate change 2021: The physical science basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge, UK: Cambridge University Press. Forthcoming 2021.
8. Kulp SA, Strauss BH. New elevation data triple estimates of global vulnerability to sea-level rise and coastal flooding. Nat Commun. 2019;10:4844. https://doi.org/10.1038/s41467-019-12808-z
9. Philip S, Kew SF, Oldenborgh GJ, Yang W, Vecchi GA, Anslow FS, et al. Rapid attribution analysis of the extraordinary heatwave on the Pacific Coast of the US and Canada June 2021. World Weather Attribution; 2021.
10. Kreienkamp F, Philip SY, Tradowsky JS, Kew SF, Lorenz P, Arrighi J, et al. Rapid attribution of heavy rainfall events leading to the severe flooding in Western Europe during July 2021. World Weather Attribution; 2021.
11. Seneviratne SI, Zhang X, Adnan M, Badi W, Dereczynski C, Di Luca A, et al. Weather and climate extreme events in a changing climate. In: Masson-Delmotte V, Zhai P, Pirani A, Connors SL, Péan C, Berger S, et al., editors. Climate change 2021: The physical science basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge, UK: Cambridge University Press. Forthcoming 2021.
12. Forster P, Storelvmo T, Armour K, Collins W, Dufresne JL, Frame D, et al. The earth's energy budget, climate feedbacks, and climate sensitivity. In: Masson-Delmotte V, Zhai P, Pirani A, Connors SL, Péan C, Berger S, et al., editors. Climate change 2021: The physical science basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge, UK: Cambridge University Press. Forthcoming 2021.
13. Engelbrecht FA, 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(8), Art. #085004. https://doi.org/10.1088/1748-9326/10/8/085004
14. Hoegh-Guldberg O, Jacob D, Taylor M, Bindi M, Brown S, Camilloni I, et al. Impacts of 1.5ºC global warming on natural and human systems. In: Masson-Delmotte V, Zhai P, Pörtner H-O, Roberts D, Skea J, Shukla PR, et al., editors. Global warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty. IPCC; 2018.

15. Archer ERM. Learning from South Africa's recent summer rainfall droughts: How might we think differently about response? Area. 2019;51:603–608. https://doi.org/10.1111/area.12547

16. L'Heureux ML, Takahashi K, Watkins AB, Barnston AG, Becker EJ, Di Liberto TE, et al. Observing and predicting the 2015/16 El Niño. Bull Am Meteorol Soc. 2017;98:1363–1382. https://doi.org/10.1175/BAMS-D-16-0009.1

17. Burls NJ, Blamey RC, Cash BA, Swenson ET, al Fahad A, Bopape M-JM, et al. The Cape Town "Day Zero" drought and Hadley cell expansion. NPJ Clim Atmos Sci. 2019;2, Art. #27. https://doi.org/10.1038/s41612-019-0084-6

18. Otto FEL, Wolski P, Lehner F, Tebaldi C, van Oldenborgh GJ, Hogesteeger S, et al. Anthropogenic influence on the drivers of the Western Cape drought 2015–2017. Environ Res Lett. 2018;13(12), Art. #124010. https://doi.org/10.1088/1748-9326/aae9f9

19. Garland RM, Matooane M, Engelbrecht FA, Bopape M-JM, Landman WA, Naidoo M, et al. Regional projections of extreme apparent temperature days in Africa and the related potential risk to human health. Int J Env Res Public Health. 2015;12:12577–12604. https://doi.org/10.3390/ijerph121012577.

20. Fitchett JM. Recent emergence of CAT5 tropical cyclones in the South Indian Ocean. S Afr J Sci. 2018;114(11/12), Art. #4426. https://doi.org/10.17159/sajs.2018/4426

21. Arias PA, Bellouin N, Coppola E, Jones RG, Krinner G, Marotzke J, et al. Technical summary. In: Masson-Delmotte V, Zhai P, Pirani A, Connors SL, Péan C, Berger S, et al., editors. Climate change 2021: The physical science basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge, UK: Cambridge University Press. Forthcoming 2021.

22. Gregor L, Lebehot AD, Kok S, Monteiro PMS. A comparative assessment of the uncertainties of global surface ocean $CO_2$ estimates using a machine-learning ensemble (CSIR-ML6 version 2019a) – have we hit the wall? Geosci Model Dev. 2019;12:5113–5136. https://doi.org/10.5194/gmd-12-5113-2019

23. Zickfeld K, MacDougall AH, Matthews HD. On the proportionality between global temperature change and cumulative $CO_2$ emissions during periods of net negative $CO_2$ emissions. Environ Res Lett. 2016;11(5), Art. #055006. https://doi.org/10.1088/1748-9326/11/5/055006

24. Mauritsen T, Pincus R. Committed warming inferred from observations. Nat Clim Change. 2017;7:652–655. https://doi.org/10.1038/nclimate3357

**AUTHOR:**
Jeffrey Mahachi[1]

**AFFILIATION:**
[1]School of Civil Engineering and the Built Environment, University of Johannesburg, Johannesburg, South Africa

**CORRESPONDENCE TO:**
Jeffrey Mahachi

**EMAIL:**
jmahachi@uj.ac.za

**HOW TO CITE:**
Mahachi J. Innovative building technologies 4.0: Fast-tracking housing delivery through 3D printing. S Afr J Sci. 2021;117(11/12), Art. #12344. https://doi.org/10.17159/sajs.2021/12344

**ARTICLE INCLUDES:**
☐ Peer review
☐ Supplementary material

**KEYWORDS:**
construction 3D printing, innovative building technologies, life cycle costing, social acceptability

**PUBLISHED:**
29 November 2021

# Innovative building technologies 4.0: Fast-tracking housing delivery through 3D printing

A house fulfils a fundamental need for human habitation. Acquisition of a house (through purchase or construction) is probably one of the greatest expenses for most people in South Africa. However, the complex South African housing industry has two markets: one market financed by the private sector and the other subsidised by the government. Despite several measures put in place by the government, the housing backlog in South Africa is still unacceptably high, at more than 2.3 million houses.[1]

Furthermore, in South Africa, there have been challenges of limited uptake of innovative building technologies in house construction. Fairclough[2] and Burger[3] have noted that innovations have changed how homes are made in many countries, their performance, affordability, and functionality. Although the South African regulatory environment is not prescriptive in the materials and products used in building construction, there has been a slow uptake of innovative building products compared to that in other countries. In this article, the term 'innovative building products' refers to any non-conventional building products that have been assessed and certified by Agrèment South Africa[4] (www.agrement.co.za). There are no South African National Standards to assess the performance of these products. The use of innovative building products in South Africa has important economic ramifications, including eradicating the housing backlog, providing better-quality housing and construction products, and possibly reducing the life cycle cost of the houses.

South Africa can deliver more than 160 000 houses and 80 000 houses per year in the government subsidised and private sectors, respectively, as evidenced in the 2008/2009 financial years shown in Figure 1. Since 2009, the delivery of government subsidised houses has been dropping at an alarming rate, indicating serious intervention required by government and private developers in the home-built environment. This, therefore, requires a change: an exploration of how innovation, in its broad context, can be utilised to examine the structure, characteristics, and technologies available to accelerate the delivery of houses.

A general reluctance by the construction industry to embrace technological advancement has meant that productivity is low, outdated, and lacking in dynamism and creativity. There are various contributory factors. For example, there is an insufficient collaboration between technology suppliers and contractors, inadequate knowledge transfer from one project to the next, fear and anxiety by built environment professionals to explore innovative ideas and solutions, and misperceptions on cost and acceptability of the technologies. However, the construction industry is well positioned to refine its business-as-usual productivity and efficiency models and embrace technological advances such as building information modelling, 3D printing, and augmented reality.[5,6] This article explores the potential of using 3D printing technologies to fast-track the delivery of quality houses in South Africa.

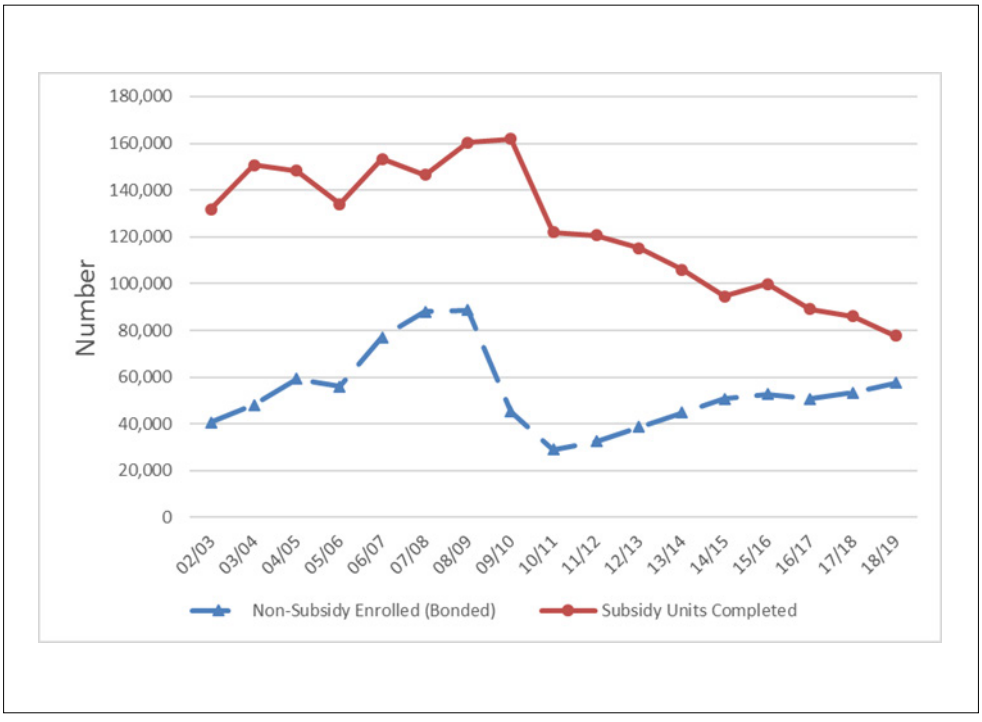



**Figure 1:** Delivery of houses (adapted from Mahachi[1]).

## Significance of the research

Three-dimensional (3D) construction printing is an innovative technology that will significantly alter the way housing will be delivered in South Africa. However, the following questions and issues need to be addressed to support the effective delivery of sustainable human settlements:

- Is 3D printing technology an appropriate future technology in a developing country such as South Africa? Is it a transformative technology that could revolutionise the house construction industry and level the playing field?
- Do 3D printing technologies offer more cost-effective products than conventional 'brick and mortar' construction?
- Conventional construction offers houses a minimum design life of 50 years. What is the life span of a 3D printed house?
- Do 3D printed houses lower operating costs (life cycle costs) over the life span of the houses compared to conventional construction? Are housing beneficiaries and developers interested in reducing operating costs in the long term, or are they only somewhat interested in short-term cost savings? Considering the current and possible future energy crisis, do housing beneficiaries appreciate the present value of energy savings over the design life of the houses?
- How easy is it to maintain and re-model a 3D printed house, and will the products be locally available during the design life of the house?
- The National Home Builders Registration Council (NHBRC)[7] is a state entity established in terms of the *Housing Consumers Protection Measures Act (Act 95 of 1998, as amended)* with a mandate to protect housing consumers through the provision of a 5-year structural warranty. If a structural defect occurs within the warranty period, the home builder must rectify the defect. However, if the builder is unavailable or fails to rectify, the NHBRC[7] is expected to make good the structural defect. Therefore, do the NHBRC[7] and its remedial contractors have the capacity and capability to remedy a 3D printed house to its original state without further compromising its structural integrity?
- With government moving towards quick economic recovery, is 3D printing the right technological solution to assist with quick delivery of houses (permanent and temporary)? Is investing in 3D construction printing technology sustainable, and is there a willingness by construction companies to invest in long-term innovation?
- Small- and medium-sized contractors are now dominating the low-income house construction – is there adequate capacity and resources to invest in 3D construction printing?

## Innovation in housing

Before exploring the potential of 3D construction printing, it is necessary to review the building regulatory environment. Performance-based building standards are concerned with what a building product is required to do, rather than with how it is done. Extensive work in this area has been reported by Fairclough[2], Foliente[8] and Hartkopf et al.[9]. The International Council for Research and Innovation in *Building* and Construction (CIB)[10] has also adopted this approach in their definition of a performance-based building. The South African building regulatory framework is also based on a similar performance-based approach. The framework provides for all buildings to comply with the National Building Regulations[11] and *Building Standards Act (Act No. 103 of 1977)*, using deemed-to-satisfy rules stipulated in South African National Standards, rational assessments, and designs by a competent person or performance assessments.

Agrèment SA[4] is a state entity established in terms of the *Agrèment SA Act (Act 11 of 2015)*, with a mandate to undertake performance assessments of construction products for structural strength and stability, fire, thermal and energy, acoustics, and durability. Upon satisfying all the performance requirements, Agrèment SA[4] issues a certificate of 'fit-for-purpose', which summarises the product's expected performance and assures that the product complies with the National Building Regulations[11].

Hartkopf et al.[9] emphasised a connection between performance-based standards and innovation. The performance-based approach in the building industry encourages innovation, allows for more competition, and supports cost-effective building. As such, the South African regulatory environment encourages and promotes innovation in construction. On this basis, this article considers innovative products as those products that have been assessed and certified by Agrèment SA[4] as meeting the performance requirements.

## Prospects of 3D construction printing

### *4IR: 3D printing technology*

It has been highlighted that one of the challenges that South Africa faces is the need to provide adequate and affordable housing and accommodation to eradicate the ever-increasing housing backlog. Some of the reasons for the slow pace of delivery include the high cost of construction methods and construction materials, unavailability of raw materials, inexperienced building contractors, and the lack of understanding and appreciation of advances in building technologies. Thus, it has become imperative to evolve a solution to the costliness of building and construction materials and possibly develop viable, cheaper, alternative materials and construction methods that embrace the goals of the Fourth Industrial Revolution (4IR).

The use of 3D printing technologies is a possible solution to provide a cost-saving and fast construction method. Although 3D printing of houses has only started to gain traction in the last few years, the technology was developed in the 1980s by Charles W Hull. Hull patented the first commercial 3D printer or stereolithographic machine in 1986. This machine functioned by having several layers of liquid ultraviolet (UV)-cured resin, one on top of the other, and then using a UV laser to trace and solidify a pattern, which in turn caused each successive layer to adhere to the previous layer. After receiving the first patent, Hull started the company 3D Systems, which commercialised the original rapid prototyping systems for CAD (Computer Aided Design) software.

In recent years, 3D construction printers use a chemically altered concrete mix pumped through a concrete extruder/nozzle controlled in three dimensions.[6] This extruder is controlled by a computerised system and builds the structure layer by layer. Therefore, the key components are (1) a concrete pump, (2) an extruder, and software. The construction process using this technology is thus mainly automated and requires minimal labour.

Cement mortar and concrete are composite materials and are the most widely used materials in the construction industry. The primary constituents of mortar and concrete are cement, a fine aggregate (sand), and water. However, these constituent materials have high production costs. In addition, the materials also have a negative impact on the environment during production. For example, Portland cement is obtained from cement clinker produced by heating powdered limestone and clay at a very high temperature. The production of 1 ton of cement is accompanied by the release of 1 ton of carbon dioxide into the atmosphere. Even the quarrying operation involved in the production of quarry dust is energy intensive. In other words, the combination of high-embodied energy and overall environmental recklessness inherent in the production of cement mortar and concrete is an essential consideration in sustainable development. Any measure aimed at reducing the embodied energy mitigates the harm to the environment and reduces the unit cost of concrete/mortar production. Such a measure is, therefore, a sustainable and environmentally friendly alternative to cement and aggregates.

The highlighted problems with traditional aggregates and Portland cement for cement mortar and concrete production necessitate research for alternative, cost-effective binding materials that can partially or wholly replace the traditional construction materials. However, vast quantities of construction, demolition, industrial and agricultural wastes

are generated from the activities related to infrastructure development. Due to these huge quantities of waste, disposal is a significant problem in South Africa and worldwide. Hence, instead of landfilling or burning this waste, which may contribute to gas emissions, it could be used as a potential substitute for natural aggregate and cement as a sustainable option for 3D printing production. This will not only convert waste to wealth but also reduce its impact on the environment. However, due consideration needs to be taken when using waste materials to ensure a high-quality and durable design mix. Any contamination in the waste may affect the quality and performance of the printed product.

There are two commonly used systems for 3D construction printer designs, i.e. the gantry-based systems (e.g. COBOD, www.cobod.com) and robotic arm type systems (e.g. Apis-cor, www.apis-cor.com). Both systems have their advantages and disadvantages, depending on the purpose of their applications. The main specifications of a 3D printer are maximum printable area (), extrusion rate (), print speed, and nozzle size (mm). Figure 2 shows an example of a 3D printed walling system and Figure 3 shows a completed 3D printed house.

### *Potentials for job creation*

As mentioned earlier, the government has attempted to address the housing crisis through the scaled delivery of subsidised housing for low-income households. The government policy makes provision for subsidy grants to specific categories of people.


*Source: www.cybe.eu/cases*

**Figure 2:** Example of 3D construction printing.

*Source: www.cybe.eu/cases*

**Figure 3:** Completed 3D printed house by CYBE.

The Department of Human Settlements launched the Breaking New Ground policy (www.dhs.gov.za) with a vision to:

- accelerate the delivery of housing as a critical strategy for poverty alleviation;
- utilising provision of housing as a significant job creation strategy; and
- leveraging growth in the economy; etc.

A critical question that needs to be answered: is 3D printing of houses going to create job opportunities, as it is perceived that only a few skilled and semi-skilled workers are required to operate the technology during construction? To effectively answer the question, the following needs to be taken into consideration:

- Resources and skills required during construction. Construction of the walling system is possibly about 30% of a completed house. Labour is still required to excavate and prepare platforms, roofing, carpentry, electrical installation, etc. Therefore the role of the community and opportunities for job creation may not necessarily be eroded by using a 3D printer to print the walls.
- Value chain of construction. The value chain for the physical construction of a house begins with identifying materials (e.g. waste), material characterisation and preparation, logistics and delivery to site, construction, and demolition/re-usage. Thus, several job opportunities will be created in the value chain and job creation should be viewed from the perspective of the total value chain cycle and not only the physical construction part.
- 3D printing may be perceived by the youth to be a 'smart' technology. Currently the industry is perceived to be 'dirty', 'dangerous' and 'disorganised'. Hence 3D printing technology has the potential to attract the youth and women to the currently male-dominated industry. With the advancement of technologies, 3D printing will provide a platform for creativity and entrepreneurship development in the house construction industry.

### *Social acceptability of 3D printed houses*

One of the inhibitors in adopting any innovative technology is the challenge of the beneficiaries' social acceptability of the product. However, 3D construction printing technology promises to provide solutions to some of the challenges, notably:

- It is a technology that provides a strong sound structure without the 'knock-on' effects associated with some of the innovative building technologies.
- The technology allows the designers to be creative and produce complex and yet aesthetically pleasing structures (houses) instead of the current 'match-box' houses that have dominated the low-income housing industry. Beneficiaries will therefore have a wider choice of customised housing typologies.

Despite this, there are also challenges that 3D construction printing technology may present for beneficiaries. Challenges include the maintenance of the house should the structure experience any structural distress and re-modelling (additions and alterations) if required. Addressing these challenges may require a combination of conventional and non-conventional techniques and the associated training of the beneficiaries; further research in this area is thus required.

### *Life cycle costing*

Another area of concern in implementing innovative building technologies, including 3D printing, is the cost of construction. Like any new technology, the cost of introducing a new technology is always high due to the required initial high capital outlay. With more uptake and usage of the technology, the costs will decrease, thereby making the product more competitive. Additionally, the cost of 3D printing a house should not be assessed from the construction costs alone. Many advantages, with indirect cost implications, come with the use of 3D printing, which will have the effect of reducing costs if a life cycle cost approach is adopted. The indirect costs considered include:

- faster delivery times;
- usage of waste materials;
- reduction in carbon emissions;
- reduction in energy consumption; and
- reduction of waste on-site.

If all the above factors are considered in the life cycle cost model, it is possible that the 3D construction printing of houses will be much cheaper than the conventional construction.

## Benefits of 3D construction printing

Using 3D printing in house construction promises many benefits to the South African housing construction industry, particularly where mass-scale house customisation is required. These can be summarised as follows:

- 3D printing offers high precision and different, complex types of typologies for the end-user. The material mix design offered is consistent, and the integrity of the structure is 'lab-based', giving a structure with the desired structural performance requirements and durability. However, it should be noted that stringent on-site quality assurance is still required to ensure a durable, uniform product.
- Material quantities required for the house construction are controlled and mixed in the right proportions with limited, if any, waste materials.
- The delivery rate is constant (with a possible 24-h production if required) and yet the same quality of production is maintained. Speed of construction has thus a potential of delivering houses much faster.
- As the houses are printed on-site, the logistics and travelling costs are reduced.
- 3D printing house construction has the potential to attract youth and women into the industry.

However, despite the benefits highlighted above, the following considerations would still need to be investigated and addressed:

- Education of professionals, mainly architects and engineers, is required to promote and adopt 3D construction printing technologies. Such education needs to start from tertiary institutions and continue through Continuous Professional Development.
- Effective collaboration is needed between the 3D printing contractor and the professional team.
- High capital outlay is needed for the 3D construction printers.
- Manufacture, of 3D printers that can be used in rugged geographic topographies, particularly in the rural areas, is needed.
- Effective installation of services and rebars (where required) is needed.
- Perceptions of beneficiaries, owners, and government on the performance of houses built through innovative building technologies in general, and 3D construction printing in particular, need to be taken into account.

## Recommendations and conclusions

The need to eradicate the housing backlog requires the government of South Africa and the private sector to partner and promote effective and innovative ways of delivering housing. The nature of innovation and the benefits of 3D printing technologies in the complex government housing value chain opens opportunities for mass-scale customisation

of houses in South Africa. 3D construction printing has not yet been explored in South Africa, although this technology has been tried and tested internationally.

The conventional 'brick and mortar' home building industry is still highly competitive, although fragmented. It requires low capital and thus makes it easy for new companies to enter and exit the market. As a result, the profit margins are meagre, particularly in the government subsidised housing market. Contractors are reluctant to invest, learn and install any new innovative products.

The government offers limited support to promote construction innovation through technical, finance, or preferential procurement. This is despite government initiatives that go as far back as 2013 when Cabinet Lekgotla resolved to use innovative building technologies to construct social infrastructure. At that time, the Cabinet resolved to set a target of 60% of the specific building types to be constructed from innovative building technologies. To date, no significant investment has been made using innovative building technologies.

To assist the government in the fast delivery of houses using 3D printing technology, further work is still required in the following areas:

- utilisation of waste materials as a cost-effective mix design;
- construction costs vs life cycle costs;
- entrepreneurial development in the value chain of housing delivery as a means to create job opportunities;
- social acceptability of the technology; and
- policy changes to support the implementation of innovative building technologies and 3D construction printing of houses.

The above would need to be verified through comprehensive pragmatic research and practical implementation of the technologies on the ground.

## References

1. Mahachi J. Development of a construction quality assessment tool for houses in South Africa. Acta Structilia. 2021;28(1):91–116. https://doi.org/10.18820/24150487/as28i1.4
2. Fairclough J. Rethinking construction innovation and research: A review of government R&D policies and practices. London: Department of Trade and Industry/Department of Transport, Local Government, England; 2002.
3. Burger S. Innovative building technologies can aid delivery of social infrastructure. Johannesburg: Creamer Media, 2014.
4. Agrèment South Africa Act No. 11 of 2015, South Africa. Government Gazette 41186, 20 October 2017.
5. World Economic Forum (WEF). Shaping the future of construction: A breakthrough in mindset and technology. Geneva. WEF; 2016.
6. Hagar I, Golonka A, Putanowicz R. 3D Printing of buildings and building components as the future of sustainable construction? Procedia Eng. 2016;151:292–299. https://doi.org/10.1016/j.proeng.2016.07.357
7. National Home Builders Registration Council (NHBRC) [homepage on the Internet]. No date [cited 2021 May 15]. Available from: www.nhbrc.org.za
8. Foliente G. Developments in performance-based building codes and standards. For Prod J. 2000;50(7/8):12–21.
9. Hartkopf V, Loftness V, Mill P. The concept of total building performance and building diagnostics. In: Davis G, editor. Building performance: Function, preservation and rehabilitation. ASTM STP 901. West Conshohocken, PA: ASTM International; 1986, p. 5–22. https://doi.org/10.1520/STP23009S
10. CIB. Performance based building: First international state-of-the-art report. Rotterdam: CIB Development Foundation, PeBBu Thematic Network; 2003.
11. National Building Regulations and Building Standards Act No. 103 of 1977, as amended, South Africa.

**AUTHORS:**
Donrich Thaldar[1]
Meshandren Naidoo[1]

**AFFILIATION:**
[1]School of Law, University of KwaZulu-Natal, Durban, South Africa

**CORRESPONDENCE TO:**
Donrich Thaldar

**EMAIL:**
ThaldarD@ukzn.ac.za

**HOW TO CITE:**
Thaldar D, Naidoo M. AI inventorship: The right decision? S Afr J Sci. 2021;117(11/12), Art. #12509. https://doi.org/10.17159/sajs.2021/12509

**ARTICLE INCLUDES:**
☐ Peer review
☐ Supplementary material

**KEYWORDS:**
artificial intelligence, patent law, intellectual property, inventorship, innovation

**PUBLISHED:**
29 November 2021

# AI inventorship: The right decision?

How should our patent system deal with inventions by artificial intelligence (AI) systems? The first hurdle to patenting an invention by an AI system is for the AI system to qualify as an inventor in our current patent law. South Africa's *Patents Act 57 of 1978* refers to an inventor as 'him'. The *Interpretation Act 33 of 1957* provides that reference to the male gender includes the female. However, does 'him' include an AI inventor (as opposed to a human inventor)? If one adopts a literal reading, and assumes that an AI system cannot be referred to as 'him', the answer would be no. But is it the *purpose* of the *Patents Act* to only provide patent protection to certain kinds of inventions, namely *human* inventions, and not *AI* inventions?

This question rapidly entered the realm of reality with the news that South Africa's Patent Office (SAPO), which runs under the auspices of the Companies and Intellectual Property Commission, granted a patent for which the inventor is an AI system.[1-3] The AI system, called DABUS (an acronym for 'device for the autonomous bootstrapping of unified sentience'), was created by American AI entrepreneur Dr Stephen Thaler. DABUS invented a new food container, which was the subject of the patent application.[4] In the patent application, Thaler is indicated as the patent owner, and DABUS as the sole inventor. This patent application was not restricted to South Africa. Thaler also submitted the same food container patent application in various other jurisdictions. In light of the fact that DABUS's food container patent application had already been rejected by the leading patent offices of the world, namely the European Patent Office (EPO)[5] and the United States Patent and Trademark Office (USPTO)[6], SAPO's decision to grant a patent, for which DABUS is the inventor, was received with a mixture of fascination and disbelief by the intellectual property (IP) community.[1,7,8] In fact, some commentators even thought that the SAPO decision was an error, or an oversight due to South Africa's formal (but not substantive) examination system for patent applications.[7-9] In this Commentary, we suggest that the SAPO decision – whether intentional or not – was the right decision from a legal perspective.

The SAPO decision was soon to find support – although implicit and from a different jurisdiction. Just a few days after publication of the SAPO decision, the Australian Federal Court handed down a judgement that is likely to have the same effect as the SAPO decision.[9]

## The Australian decision

Thaler filed patent applications on DABUS's food container in various jurisdictions, including Australia.[10] However, the Australian Deputy Commissioner of Patents rejected the patent application[11], whereupon Thaler sought legal redress[9]. The Deputy Commissioner argued that because the *Australian Patents Act 83 of 1990* does not define the term 'inventor': (1) an AI system does not qualify as an inventor based on the ordinary meaning of the word, which requires an inventor to be a human; and (2) it would be impossible to identify a person who would thereafter be granted a patent.[9]

In a groundbreaking decision[9], the Australian Federal Court per Justice Beach set aside the Deputy Commissioner's decision and referred the patent application back to the Deputy Commissioner for reconsideration. The core of the matter was the same as that in the EPO and the USPTO, namely whether an AI system, DABUS, qualifies as an inventor for the purposes of patent law. However, the court decided contrary to those patent offices. In evaluating argument (1) above, the court found that excluding AI systems from the meaning of 'inventor' would lead to an unacceptable situation whereby any invention by an AI system would be unpatentable which would run contrary to the object of the *Australian Patents Act* – which is to, inter alia, promote technological innovation. The court held that 'inventor' need not be interpreted narrowly, but its meaning should rather evolve to meet the objects of the *Australian Patents Act*; and that including AI systems in the meaning of 'inventor' is a recognition of the reality that AI systems are in fact inventing. In dealing with argument (2), the court found that the Deputy Commissioner confused the concepts of ownership and inventorship. The inventor is not necessarily the owner, and while AI systems can be the inventor, it cannot be the owner. This is because only a person (in the legal sense) is capable of *ownership*, but the same is not applicable to *inventorship*, which only requires that the person – or object – can create a patentable invention.

However, an interesting counter-argument was presented, namely that ownership of a patent flows from the inventor; ergo the inventor must have the legal capacity to assign his or her (or its) rights in the invention and communicate this intention. As AI systems do not have legal personhood, they cannot bear rights nor assign them. Although this seems to be a strong argument, the court rejected it. The court pointed out that (at least in Australian patent law) ownership of a patent need not be acquired by assignment of rights from the inventor to the would-be owner, but can be based on any legal reason that provides a similar entitlement to ownership. Using established principles of property law, the court held that due to Thaler's ownership and control of DABUS, he would automatically be entitled to any invention by DABUS.

In a historic and well-drafted judgement, the court decided that: (1) AI systems such as DABUS are not precluded from inventorship status; (2) the objective of patent law supports the notion that AI systems can be inventors; and (3) patent ownership vests in the owner of the AI inventor.

## Why the difference between Australia, Europe and America?

We now take a step back to consider the European and US decisions on DABUS's food container patent application that preceded the South African and Australian decisions on the same patent application – and came to a different conclusion about AI inventorship. We first consider the European decision, dating from January 2020, where

the EPO highlighted that, according to the European Patent Convention, inventorship involves complementary rights and is itself a title right. As AI systems do not have legal personhood – and therefore cannot be the bearer of any of these rights – the EPO took the position that AI systems are precluded from inventorship status.[5] The problem with this argument is that it assumes that the inventor must always be able to be the bearer of rights. Consider a situation in which a human inventor dies before his or her patent application is submitted to the patent office. The right to apply for a patent now vests with the legal representative of the deceased's estate. Ownership upon grant would then vest in the deceased estate. Although the deceased *qua* inventor is not a legal person capable of being the bearer of rights any longer – it will still be required by law that the deceased is to be rightfully named as the inventor for the purposes of the application. This example clearly displays that, at the time that application is made, legal personhood is not necessary to be named as an inventor in a patent application, and ownership rights can vest in a legal person based on the operation of law other than assignment by the inventor.

Similarly, as pointed out by the Australian Federal Court, the well-established rules of property law provide a solution for AI inventorship. The court relied on the rule that the owner of a principal object is deemed to be the owner of the fruits of such object. This rule also applies in South Africa. Applied to the facts of the case, this rule entails that Thaler *qua* owner of DABUS is deemed to be the owner of DABUS's fruits, namely the food container invention.

In April 2020, the USPTO also rejected DABUS's food container patent application. It followed a strict black letter law approach to statutory interpretation, finding that the term 'inventor' only includes natural persons as the United States Code refers to pronouns such as 'himself', 'herself', and 'whomever'.[6] This stands in contrast with the purposive approach to statutory interpretation adopted by the Australian Federal Court. The USPTO also argued that central to US patent law is the mental act of conception, of which AI systems are deemed incapable.[6] The court disposed of this argument as follows:

> [W]*hat is meant by 'mental act' or 'thought'? If you simply define this as something engaged in within the human cerebral cortex, not only have you not defined any complete cognitive content but you have conveniently defined away the problem at hand.*[9]

## A broader reading of the *South African Patents Act*

The *South African Patents Act* and its regulations reveal many of the same issues that were highlighted by the EPO, the USPTO, and the Australian Deputy Commissioner in their jurisdictions. In particular, as mentioned above, the *South African Patents Act* refers to the inventor as 'him'. However, the point of departure when interpreting the *Patents Act* should be that South African law adheres to a *purposive approach* to statutory interpretation.[12] The *Patents Act* itself states rather tersely that its objective is to 'provide for the registration and granting of letters patent for inventions and for matters connected therewith'. What is the purpose of legally providing for patenting of inventions? There are various theories, justifications, and values that underlie patent law, including promoting fairness[13], justice[14], disclosure[15], reward[15], research and development[16], innovation[16], and human flourishing[17]. The Intellectual Property Policy of the Republic of South Africa Phase 1[18] (IP Policy) suggests the following purposes for IP, which includes patents:

> *Intellectual Property (IP) is an important policy instrument in promoting innovation, technology transfer, research and development (R&D), creative expression, consumer protection, industrial development and more broadly, economic growth.*[18]

We suggest that the IP Policy provides a useful articulation of the purposes served by patents, and therefore by the *Patents Act*. The fact that the IP Policy postdates the *Patents Act* is not a concern, because the purposes themselves are not time specific.

Given the reality that AI technology has reached a stage of development at which AI systems may be capable of autonomous inventing – as demonstrated by DABUS's food container patent application – we suggest that these purposes suggested in the IP Policy would be served by allowing the patenting of inventions by AI systems. Ergo, the term 'inventor' ought to be interpreted as including AI systems. Furthermore, the *Patents Act* requires that the applicant provides proof of title or authority, but does not limit it to assignment by the inventor, hence leaving open the possibility of title based on ownership of the AI inventor. We therefore suggest that the decision by the SAPO to grant a patent for DABUS's food container to Thaler, and thereby allowing for AI inventorship, was the right decision in South African law.

There are a number of important similarities between South African and Australian patent law, most pertinently that neither the *South African Patents Act* nor the *Australian Patents Act* defines 'inventor' or confines the right to apply for a patent to assignment. As such, we suggest that the rationales underlying the judgement by the Australian Federal Court can find fruitful application in South African law.

## Does this mean that AI has personhood?

Does AI inventorship point toward AI personhood (legal subjectivity)? Not necessarily. There can be both moral and economic reasons for expanding personhood to AI. From a moral perspective, if AI attains self-consciousness, this would at least entitle it to moral status aligned with the higher animal species. But for personhood – moral status on a par with humans – AI would likely need to also have the full gamut of human emotions and the ability to form relationships. From an economic perspective, it might be useful to make an AI system the bearer of its own rights and duties. Similar to a company, a (human) board of directors can manage the AI system's affairs. However, none of these are necessarily implied by allowing AI inventorship. To the extent that one views inventiveness as a human characteristic, AI now has this human characteristic (at least in the way South African law is being applied by the SAPO). That, however, does not imply that AI is fully human or is legally viewed as a person. Rather, it means that AI is a *special species of legal object that has the ability to invent*.

A legal object can have no rights or duties. As such, AI systems can be owned and disposed of like any other computer software or hardware. This, at least, is the status quo. There might arise moral or economic reasons to change this status quo in future. Importantly, although the foundational legal classification between persons (legal subjects) and things (legal objects) is binary, there is space to construct nuanced legal rules on these foundations that can provide for unique treatment of AI. To illustrate, although a dachshund is a legal object, legal subjects – including its owner – are legally prohibited from treating it inhumanely. And although a dachshund cannot inherit its owner's estate, its owner can establish a trust that can inherit the estate and that is entrusted with the dog's care. Even inanimate objects can receive special protection by the law. Examples are historic buildings or heritage sites. Similarly, the law can be developed in imaginative ways to cater for AI as part of human society.

## Concluding remarks

The SAPO made a historic decision to allow AI inventorship. It was a bold decision given the foreign precedents that were stacked against the SAPO at the time. However, it was the right decision from a legal perspective, and was soon vindicated by the Australian judgement on the same patent application. Moreover, the SAPO decision is also aligned with South African public policy on AI more broadly. The South African government has published various policy documents relating to science, technology, and innovation in light of the Fourth Industrial Revolution (4IR).[19-21] The common themes running through these documents are that: (1) South Africa is marred by resource limitations including that of skilled human capital; (2) policymakers intend to reform laws (including patent law) and create structures to take full advantage of the 4IR (including the creation of an AI institute); (3) innovation in South

Africa has been poor and sustainable innovation is key to improving the welfare of the country's people and economy; (4) technology (such as AI and cloud computing) is central to the solutions to South Africa's plans going forward.

The fact that South Africa has become the first jurisdiction in the world to allow AI inventorship should be hailed as progressive and pro-science. AI innovation has the potential to improve the human condition. Today's food container may be tomorrow's life-saving medicine.

## References

1. Patrick A. SA first to award patent recognising artificial intelligence as inventor. Times Live. 2021 July 30 [cited 2021 Jul 31]. Available from: https://www.timeslive.co.za/news/south-africa/2021-07-30-sa-first-to-award-patent-recognising-artificial-intelligence-as-inventor/
2. Smith J. South Africa issues world's first patent naming AI as inventor [webpage on the Internet]. 2021 July 29 [cited 2021 Jul 31]. Available from: https://www.mathys-squire.com/insights-and-events/news/south-africa-issues-worlds-first-patent-naming-ai-as-inventor/?utm_source=Mondaq&utm_medium=syndication&utm_campaign=LinkedIn-integration
3. Basham V. South Africa issues world's first patent listing AI as inventor. The Global Legal Post. 2021 July 28 [cited 2021 Jul 31]. Available from: https://www.globallegalpost.com/news/south-africa-issues-worlds-first-patent-listing-ai-as-inventor-161068982
4. Thaler SL. Food container and devices and methods for attracting enhanced attention. ZA2021/03242. 2021.
5. European Patent Office. Decision of 27 January 2020 on EP 18 275.
6. United States Patent and Trademark Office. Decision on petition in re application of application no.: 16/524,350. 27 April 2020.
7. Conlon E. DABUS: South Africa issues first-ever patent with AI inventor [webpage on the Internet]. 2021 July 29 [cited 2021 Jul 30]. Available from: https://www.managingip.com/article/b1sx9mh1m35rd9/dabus-south-africa-issues-first-ever-patent-with-ai-inventor
8. Swabey P. South Africa's AI patent approval could trigger innovation and abuse. Tech Monitor. 2021 July 30 [cited 2021 Jul 31]. Available from: https://techmonitor.ai/technology/ai-and-automation/south-africas-ai-patent-approval-could-trigger-innovation-and-abuse
9. Thaler v Commissioner of Patents [2021] FCA 879.
10. Thaler SL. Food container and devices and methods for attracting enhanced attention. Patent application no. 2019363177. 17 September 2019.
11. Stephen L Thaler [2021] APO 5.
12. Bato Star Fishing (Pty) Ltd v Minister of Environmental Affairs and Tourism 2004 (4) SA 490 (CC).
13. Merges RP. Justifying intellectual property. Cambridge: Harvard University Press; 2011. https://doi.org/10.4159/harvard.9780674061125
14. Ho CM. Do patents promote the progress of justice – Reflections on varied visions of justice. Loy Uni Chi LJ. 2005;36(2):469–482. https://lawecommons.luc.edu/luclj/vol36/iss2/9
15. Du Bois M. Justificatory theories for intellectual property viewed through the constitutional prism. PER. 2018;21:2–38. http://dx.doi.org/10.17159/1727-3781/2018/v21i0a2004
16. Grabowski HG, DiMasi JA, Long G. The roles of patents and research and development incentives in biopharmaceutical innovation. Health Affairs. 2015;34(2):302–310. https://doi.org/10.1377/hlthaff.2014.1047
17. Du Bois M. State use provisions for patent law, and expropriations: Some comparative law guidelines for South Africa during the Covid-19 crisis and beyond. PER. 2020;23:2–37. http://dx.doi.org/10.17159/1727-3781/2020/v23i0a8150
18. South African Department of Trade and Industry (DTI). Intellectual Property Policy of The Republic of South Africa Phase 1 2018. Pretoria: DTI; 2018.
19. South African Department of Science and Technology (DST). White Paper on Science, Technology and Innovation. Pretoria: DST; 2019. Available from: https://www.dst.gov.za/images/2019/White_paper_web_copyv1.pdf
20. South African Department of Communications and Digital Technologies. Electronic Communications Act, 2005 (Act No. 36 of 2005). Invitation to submit written submissions on the proposed National Data and Cloud Policy. Government Gazette No. 44389:306. 1 April 2021.
21. South African Department of Communications and Digital Technologies. Summary report & recommendations: Presented by the Presidential Commission on The Fourth Industrial Revolution. Government Gazette No. 43834:121. 23 October 2020.

**AUTHOR:**
Irvy M.A. Gledhill[1]

**AFFILIATION:**
[1]School of Mechanical, Industrial and Aeronautical Engineering, University of the Witwatersrand, Johannesburg, South Africa

**CORRESPONDENCE TO:**
Irvy Gledhill

**EMAIL:**
Igle.Gledhill@wits.ac.za

**HOW TO CITE:**
Gledhill IMA. Are we there yet? Academies, scientific organisations, and gender. S Afr J Sci. 2021;117(11/12), Art. #12757. https://doi.org/10.17159/sajs.2021/12757

**ARTICLE INCLUDES:**
☐ Peer review
☐ Supplementary material

**KEYWORDS:**
universality, gender transformation, apex organisations

**PUBLISHED:**
29 November 2021

# Are we there yet? Academies, scientific organisations, and gender

Is the representation of women in academies and scientific organisations improving, globally? A major new study shows that among the academies, representation of women in membership has risen from 13% (for academies surveyed in 2015) to 16% (for all academies surveyed in 2020).[1] The progress is good: the level of representation is shocking. In the words of Daya Reddy, past President of the International Science Council, '… Societies expect more diverse gender representation in science.'[2]

This report, however, is a goldmine of information on gender in leading scientific organisations. The study explores the dimensions of differences across disciplines and regional variation, and provides 10 clear recommendations.

## Global partnerships in science

The study was a collaboration of the Gender in Science, Innovation, Technology and Engineering (GenderInSITE), the InterAcademy Partnership (IAP), and the International Science Council (ISC), and was published in September 2021.[1] In these organisations, two surveys of gender representation among academies (in IAP and ISC) and scientific unions and global organisations (in ISC) have provided some answers to the perennial question of whether there is equitable representation and recognition of women in these senior levels of the scientific community.

IAP has a membership of about 140 academies of science. ISC brings together 40 international scientific unions and associations, and more than 140 regional and national academies and scientific organisations. Together, they work across natural, mathematical and computational sciences, social sciences and the humanities, medical and health sciences, and, in some cases, arts and law. They are apex organisations of science in the inclusive and general sense of the word, as defined by UNESCO.[3] These academies, scientific unions, and global organisations uphold rationality and ethics, professional integrity, and collegiality of science. In concert, they tackle the wicked problems of the planet, because they can draw on the top intellect of the global community in climate change, in pandemics, and in sustainable development. They act through providing evidence-based scientific advice to their governments and stakeholders, at global, multilateral and national levels, and through foreseeing and defining global scientific agendas. Their day-to-day work includes providing the environment for international agreement within and across sciences, for example on the names of new elements, and endorsement of the definition of the kilogram.[4,5]

The fair representation of women in these functions is essential. The principle of universality in science[6] links the free and responsible practice of science with equitable opportunities for access to science, and firmly opposes discrimination based on factors such as ethnic origin, religion, citizenship, language, political or other opinion, disability, and age, as well as sex, gender identity and sexual orientation. More recently, attention has also been focused on the intersectionality between these dimensions. This report specifically focuses on the situation of women, in terms of Sustainable Development Goal 5. At this point, it would be wise to say that the word 'we' will be used in this article for scientists and academicians of any gender, and is intended to be inclusive. Thus, 'we' – specifically including men – should benefit from improvement in our professional practices designed to foster the progression of women. Achieving that aim is a joint project.

The surveys were carried out by GenderInSITE (Gender In Science, Innovation, Technology and Engineering) – an international organisation promoting the role of women. It has a track record of finding deeper insights into science, more effective programmes, and more sustainable outcomes.

The methodology of the surveys was straightforward and simple, and facts were collected from 85 academies and 38 scientific unions and organisations. Comparisons were made for those academies included in the preceding IAP survey of 2015[7], performed by the Academy of Science of South Africa (ASSAf). This short Commentary cites only a shop window of results.

## Academies

The total number of academies surveyed was 215. The number providing usable responses was 85. For all senior academies in this new report, the representation of women as members in 2020 was 16% (Figure 1).

Figure 2 shows the distribution of women's share of membership. Among the senior academies, only Cuba, Belgium, and Venezuela host academies with more than 30% women members. In ASSAf, 27% of members are women (Figure 1); elections in 2021 are likely to have increased this number. The lowest memberships of women, below 5%, are found in academies hosted in Mongolia, Brazil (medicine), the Islamic Republic of Iran, and the Republic of Korea. Women's share of membership is at or below 10% in 19 academies (Figure 1).

Why are these fractions so low, in view of the fact that academies are top-level organisations, setting trends for the world? We might ask if the selection pool is adequate. It will immediately be noticed from Figure 2 that 10 of the young academies are near parity (45% to 55%). The South African Young Academy of Science is at the top of the list of all academies at 57%. A recommendation of the report is that senior academies can benefit by paying attention to the achievements of the younger academies. Do young academies provide a stream of bright scientists to the more elderly academies? If we compare the absolute numbers of members, in Figure 2, we observe that the

numbers of members in young academies is small enough to warrant a great deal of investment if they are to make a difference in national academy membership.

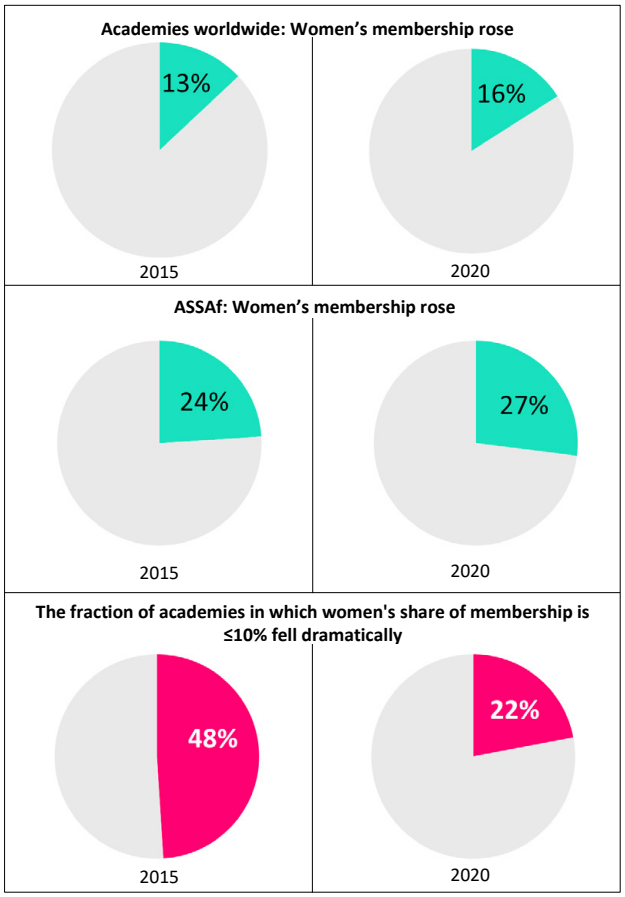



**Figure 1:** Representation of women in academy membership. Figures are calculated for participants in each survey, in 2020[1] and 2015[7].

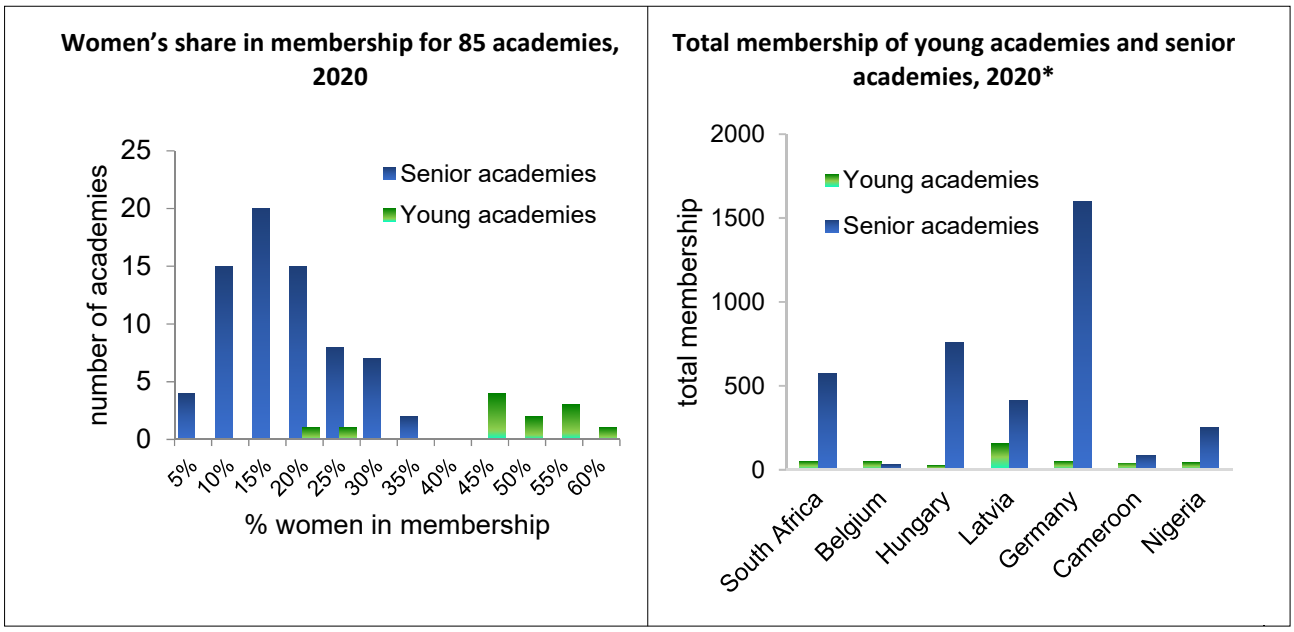
A second possible reason for large gender gaps may be inaction by academies on gender issues. Of the 215 academy members of IAP and ISC, only 85 returned unique usable responses. Then, in an inspired question in the 2020 survey, GenderInSITE enquired of participants in the first survey of 2015 whether they had read or acted upon that survey report. Only 4 of the 20 respondents indicated that they took any action to implement any of the four recommendations made in that report: collecting and analysing gender disaggregated data, publishing those data in their annual report, reporting on the gender dimensions of IAP's activities, and providing permanent organisational structures for strategic direction on gender mainstreaming (Figure 3). This could be due to their assessment that they did not need to take any new actions.

Thirdly, we note that women are lost from scientific careers at a higher rate than men. Both men and women face serious obstacles in science; both overcome barriers and make real impact. There are, however, disadvantages affecting women in almost every aspect of a career in science, as was demonstrated in the Gender Gap project sponsored by ISC, eight scientific unions, GenderInSITE, the Organization for Women in Science for the Developing World (OWSD) and UNESCO over the years 2017–2020.[8] Women report that they are treated with less respect than men, and have access to fewer facilities and opportunities. Gender gaps that may appear minor at first glance tend to compound over the years, in a process of cumulative disadvantage for women.[9]

A data-backed analysis of gender patterns in publication usefully quantified attrition rates in the production of papers. Using automated gender inference[10] on a comprehensive mathematical publication database, authorships could be tracked over time. For mathematicians, the attrition rates were about 5% higher, over a 10-year period, for women than for men.[8] There is a significant gender gap in publication, with an impact on promotion and grant success.[8]

These factors are likely to leave the academies in a position in which the selection pool contains significantly fewer women than men.

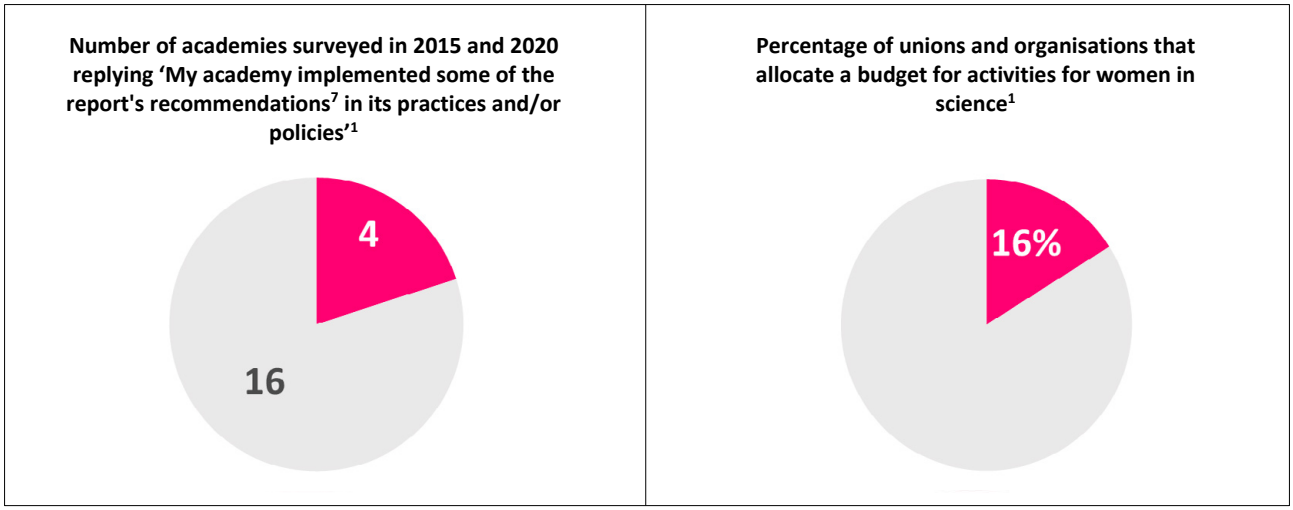
A fourth point: in academy elections, are we encountering conscious or unconscious bias (among both women and men), possibly compounded by cumbersome election mechanisms, where it is easier to vote for well-known colleagues than to make the effort to assess accomplishments? Perceptions of brilliance connected to gender are known to influence career choices[11] and might possibly influence elections.




**Some countries host more than one senior academy; only one for each country is used for illustration here.*

*Data source: 'Results of two global surveys'[1].*

**Figure 2:** Can young academies change the pipeline?




*Data source: 'Results of two global surveys'[1].*

**Figure 3:** Two indications of actions.

The Academy of Sciences in Cuba pointed out that without continued work, 'there exists the possibility of moving backward if we don't keep the activism and underlining gender issues in the internal life of the academy'[1]. How is merit judged by voting members in a population in which measures of excellence have been defined to suit a historically dominant population? For women, presentation of selected top papers is likely to be a better indicator of excellence than a count of published papers.[12]

Hope is provided by Ngila et al.[13]:

> *Science academies are well placed to contribute towards strengthening of national systems of innovation through advocating for an increased participation of girls and women in science. To successfully do so, academies would need to overcome challenges faced with regard to women's representation in their own ranks and women's resultant full participation in the activities of national science academies.*

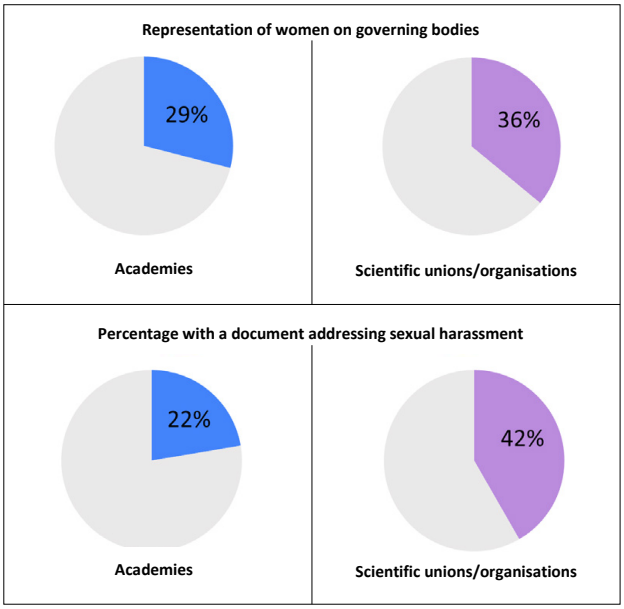
## Unions and scientific organisations

Of the 38 disciplinary unions and associations surveyed, about half have only countries as members. Membership could not be surveyed as thoroughly as among academies, and the results for disciplinary unions and scientific organisations are discussed in comparisons below. These organisations have membership and strategic mechanisms different from those of academies.

## Leadership of leadership

Academies and scientific organisations have had more success in reaching for gender parity in their leadership. Women are presidents or co-chairs of 21% of academies, and 37% of unions and scientific organisations. Representation on governing bodies is considerably higher than the membership averages (Figure 4) and the percentage of women in the governing bodies in unions and associations ranges from 67% in the social sciences to 24% for physical and natural sciences.

It is worth noting that *both* men and women in presidential or chairing positions exercise a surprisingly large degree of influence in setting a course for inclusion and diversity in strategy and membership, and eradicating systemic bias. A leader who shirks the responsibility for universality is able to bring the entire organisation to a halt, not only in terms of gender, but in terms of racial diversity, intersectionality, and the dimensions of fairness, respect, and integrity.




*Data source: 'Results of two global surveys'[1].*

**Figure 4:** Leadership and a sample of policy documentation.

## Budget

The painful experience of gender champions is that very little is accomplished without a budget. Among disciplinary unions and associations, 16% allocate a budget to implement activities related to gender equality (Figure 3). In the International Union of Pure and Applied Physics (IUPAP), the largest single allocation to a Commission or Working Group is that for Women in Physics. This ensures anchor funding for major events, and provides women with conference travel grants. 'Technology grants' were introduced in 2020 to enable women in developing countries to fund data and set up connectivity to attend virtual conferences.

## Sexual harassment

The ISC Gender Gap Global Survey[8], to which there were 32 346 respondents from 159 countries, provides clear evidence concerning sexual harassment in the careers of women. Those that indicated that they had personally experienced sexual harassment at school or work were approximately 25% of women and approximately 4% of men, with some variation across disciplines and regions. Sexual harassment cannot be tolerated. Major scientific organisations do have a role to play,

through their policies, statements, ethics, and their expectations of their member organisations. The apex organisations have the power to set the example for the scientific community.

However, few academies, but more unions and organisations, have a document that addresses sexual harassment in the workplace (Figure 4). These policies sometimes include a Code of Conduct, or a policy for conferences, but action in a case is usually a harrowing process for all involved. Building strong culture and science identity within a discipline is an essential task of academies and unions to prevent unacceptable behaviour and provide a safe, welcoming, and inclusive environment.

## Disciplines and regions

The report provides a useful comparison across disciplines and regions. It is not unexpected that the largest representation of women is found in biological and social sciences, with computer and mathematical sciences trailing – but this fact in itself should lead us to investigate the contrasts in environments and career choices, to learn what is working so effectively. As observed in the participant feedback in the report, 'Global comparisons present the current status, and the relevant evidence has the potential to prompt both established and new academies into action'[1]. The unions emphasised that regional action is essential.

## Working in concert

Survey methodologies fit together in a jigsaw. The UNESCO SAGA (STEM and Gender Advancement) structure for measuring gender equality[14] is an outstanding framework and similar to that used in this report. Other methodologies provide interlocking reading matter. The Global Survey of mathematical, computing, and natural scientists used a snowball sampling method through ISC disciplinary unions and associations to shed light on *what* obstacles and successes are encountered by both men and women.[8] The data-backed study of patterns in publication provides profound insight through automated gender inference for millions of papers and preprints.[8,10] Conceptual structure, based on the SAGA Gender Objectives List[15], was developed and tested, and was used to research and produce a database of evaluated interventions[8].

Many academies and organisations have cooperative links with OWSD, TWAS (The World Academy of Science), UNESCO, GenderInSITE, and their regional umbrellas such as the Network of African Science Academies (NASAC). Many celebrate 11 February, the International Day of Women and Girls in Science. The Standing Committee on Gender Equality in Science was formed in 2021, and works across scientific unions and associations.

## Foresight

Why is it that, when initiatives to empower women in science come to an end, the situation often snaps back to the way it was originally? I have often asked the question of my colleagues in a spirit of inquiry. The reply which appears to hold the key is: 'Because the underlying culture has not changed.' We have had a startling illustration of this answer with the effect of the pandemic on the careers of academic women in South Africa[16], and we may well be facing disastrous consequences for equality in years to come.

This study[16] shows that, during the COVID-19 pandemic, a third of women surveyed reported making no progress towards a significant academic product, and 95% with toddlers indicated that childcare had a high impact on their work. Of competing household chores with highest impact, 44% indicated childcare and 43% schoolwork. Universities are called on to acknowledge this potential problem, and to adjust the timelines in the appointment and advancement of female academics.[16]

## Are we there yet?

No, we are not there yet. There has, however, been encouraging, measurable improvement. Some organisations are attaining parity in membership, notably young academies. Their practices lead the way: recommendation 9 suggests, for example, that senior academies would do well to learn from the practices of young academies.

It is discouraging to note that 19 academies have women's share of membership at or below 10%, and that of 215 academies, only 85 responded.[1] It is a not-unexpected observation that useful reports are not read, still more that no action is taken – unless, of course, no action was necessary, a proposition that awaits investigation within the data that have been presented.

The President of ISC, Peter Gluckman, comments:

> *This survey confirms what many of us suspected. The number of science organisations systematically monitoring progress on gender equality remains low and must be extended. The report also highlights insufficient knowledge about action on gender equality in science institutions around the world. Large umbrella organisations like the ISC and the IAP have a responsibility to ensure that we, as a global scientific community, assess progress and exchange of information about policies that work in different contexts. I hope that what we have now started with our membership will continue and inspire other organisations.*

## Final word

As a source of potential actions, this report is outstanding. It provides a route map through which organisations and academies can locate and refer to each other's documents and charters – the appendices are particularly rich and useful in this respect for any organisation that is seeking tested interventions. It is a good handbook for comparison between disciplines, regions and organisational structures. Reporting across countries and regions prompts introspection and possibly action. This report should be read, and its recommendations (see Box) should be considered at a strategic level.

The gender lens should be applied by all leaders of academies, scientific unions, and global organisations, to fulfil their mandate and take a stand for humanity in science, as well as science for humanity. The anticipated setback associated with COVID-19 challenges in the careers of women will need both acknowledgement and action.

**Recommendations**

Each of these thumbnail headings is fully explained in the source report[1].

1. Extension of survey
2. Analysis of gender-related organisational policy, structure and actions
3. Development of a central repository
4. Incorporation of regional considerations
5. Advancing women to leadership positions
6. Consideration of diversity and inclusivity
7. Analysis of discipline-based gender transformation
8. Establishment of monitoring and evaluation frameworks
9. Identification of lessons from young academies
10. Shift from a focus on 'numbers' to institutional and knowledge transformation

A recommendation within the present Commentary: read the report.

## Acknowledgement

Data from the 'Results of two global surveys'[1] were published under a CC BY 4.0 licence and have been used to construct the figures in this Commentary.

## Competing interests

I have no competing interests to declare.

## References

1. GenderInSITE, InterAcademy Partnership (IAP), International Science Council (ISC). Gender equality in science: Inclusion and participation of women in global science organizations: Results of two global surveys [document on the Internet]. c2021 [cited 2021 Oct 20]. Available from: https://www.interacademies.org/publication/gender-equality-science-inclusion-and-participation-women-global-science-organizations
2. International Science Council (ISC). Media release: Gender equality in science: Inclusion and participation of women in global science organizations [webpage on the Internet]. c2021 [cited 2021 Oct 29]. Available from: https://council.science/current/press/gender-equality-in-science-inclusion-and-participation-of-women-in-global-science-organizations/
3. United Nations Educational, Scientific and Cultural Organisation (UNESCO). Final report on the draft text of the UNESCO Recommendation on Open Science [document on the Internet]. c2021 [cited 2021 Oct 20]. Available from: https://unesdoc.unesco.org/ark:/48223/pf0000376130?posInSet=6&queryId=854724a2-cf19-4d1f-8b54-836ba837d5b5
4. Chatt J. Recommendations for the naming of elements of atomic numbers greater than 100. Pure Appl Chem. 1979;51:381–384. https://doi.org/10.1351/pac197951020381
5. Bureau International des Poids et Mesures (BIPM). The International System of Units (SI). 9th ed. Paris: BIPM; 2019. Available from: https://www.bipm.org/documents/20126/41483022/SI-Brochure-9-EN.pdf
6. International Council for Science (ICSU). Freedom, responsibility and internationality of science [document on the Internet]. c2014 [cited 2021 Oct 20]. Available from: https://council.science/wp-content/uploads/2017/04/CFRS-brochure-2014.pdf
7. Academy of Science of South Africa (ASSAf). Women for science: Inclusion and participation in academies of science. Pretoria: ASSAf; 2016. http://hdl.handle.net/20.500.11911/7
8. Roy M-F, Guillopé C, editors. A global approach to the gender gap in mathematical, computing, and natural sciences: How to measure it, how to reduce it? Berlin: International Mathematical Union; 2020. Available from: https://zenodo.org/record/3882609#.YXa5bBpBw2w
9. Valian V. Why so slow? The advancement of women. Cambridge, MA: MIT Press; 1997.
10. Mihaljević H, Tullney M, Santamaría L, Steinfeldt C. Reflections on gender analyses of bibliographic corpora. Front Big Data. 2019;2(29):1–6. https://doi.org/10.3389/fdata.2019.00029
11. Leslie S-J, Cimpian A, Meyer M, Freeland T. Expectations of brilliance underlie gender distributions across academic disciplines. Science. 2015;347:262–265. https://doi.org/10.1126/science.1261375
12. Declaration on Research Assessment (DORA). San Francisco Declaration on Research Assessment [webpage on the Internet]. No date [cited 2021 Oct 20]. Available from: https://sfdora.org/read/
13. Ngila D, Boshoff N, Henry F, Diab R, Malcom S, Thomson J. Women's representation in national science academies: an unsettling narrative. S Afr J Sci. 2017;113(7/8), Art. #2017-0050. https://doi.org/10.17159/sajs.2017/20170050
14. United Nations Educational, Scientific and Cultural Organisation (UNESCO). Measuring gender equality in science and engineering: The SAGA toolkit. SAGA Working paper 2. Paris: UNESCO; 2017. Available from: http://uis.unesco.org/sites/default/files/documents/saga-toolkit-wp2-2017-en.pdf
15. United Nations Educational, Scientific and Cultural Organisation (UNESCO). The SAGA Science, Technology and Innovation Gender Objectives List (STI GOL). SAGA Working paper 1. Paris: UNESCO; 2016. Available from: https://unesdoc.unesco.org/ark:/48223/pf0000245006
16. Walters C, Mehl GG, Piraino P, Jansen JD, Kriger S. The impact of the pandemic-enforced lockdown on the scholarly productivity of women academics in South Africa. Res Policy. 2022;51(1), Art. #104403. https://doi.org/10.1016/j.respol.2021.104403

**AUTHORS:**
Elmar van Rooyen[1]
Trudy Paap[2,3]
Wilhelm de Beer[2,3]
Garyn Townsend[4]
Shawn Fell[2,3]
Wilma J. Nel[2,3]
Seamus Morgan[5]
Martin Hill[4]
Allan Gonzalez[6]
Francois Roets[1]

**AFFILIATIONS:**
[1]Department of Conservation Ecology and Entomology, Stellenbosch University, Stellenbosch, South Africa
[2]Department of Biochemistry, Genetics and Microbiology, University of Pretoria, Pretoria, South Africa
[3]Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Pretoria, South Africa
[4]Centre for Biological Control, Department of Zoology and Entomology, Rhodes University, Makhanda, South Africa
[5]Department of Genetics, Stellenbosch University, Stellenbosch, South Africa
[6]School of Forest, Fisheries and Geomatics Sciences, University of Florida, Gainesville, Florida, USA

**CORRESPONDENCE TO:**
Francois Roets

**EMAIL:**
fr@sun.ac.za

**DATES:**
**Received:** 22 Feb. 2021
**Revised:** 22 June 2021
**Accepted:** 28 June 2021
**Published:** 29 Nov. 2021

**HOW TO CITE:**
Van Rooyen E, Paap T, De Beer W, Townsend G, Fell S, Nel WJ, et al. The polyphagous shot hole borer beetle: Current status of a perfect invader in South Africa. S Afr J Sci. 2021;117(11/12), Art. #9736. https://doi.org/10.17159/sajs.2021/9736

**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:**
distribution, *Euwallacea fornicatus*, *Fusarium euwallaceae*, host range, invasion biology

**FUNDING:**
South African Department of Forestry, Fisheries and the Environment, University of Pretoria

# The polyphagous shot hole borer beetle: Current status of a perfect invader in South Africa

The polyphagous shot hole borer (PSHB) beetle is a recent invader in South Africa. Together with its fungal symbiont, *Fusarium euwallaceae,* it can rapidly kill highly susceptible host plants. Its impact is most profound in urban areas, but it has also been found infesting important forestry, agricultural crop and native species. Since its first detection in 2012, PSHB has spread to all but one province in the country. The beetle–fungus complex has several biological traits that enhance its anthropogenically mediated dispersal, establishment and survival in novel environments – factors that have likely facilitated its rapid spread across the country. We review the history of the PSHB invasion in South Africa, its taxonomic status and the reasons for its rapid spread. We highlight its potential impact and challenges for its management. Finally, we provide an updated distribution map and list of confirmed host plants in South Africa. Of the 130 plant species identified as hosts, 48 of these (19 indigenous and 29 introduced) are reproductive hosts able to maintain breeding PSHB populations. These reproductive hosts may succumb to beetle infestations and act as 'pest-amplifiers'. The economic impact on urban forests, plantation forestry and agricultural crops may be severe, but the ecological impact of PSHB invasion in native ecosystems should not be underestimated.

**Significance:**

- We provide an updated host list and distribution map for South Africa of the globally significant tree pest, the polyphagous shot hole borer (PSHB, *Euwallacea fornicatus*). The South African PSHB invasion represents the largest outbreak of this pest in its global invaded range. PSHB was confirmed to infest 130 plant species in urban, agricultural, and native ecosystems in South Africa, including 44 previously unreported hosts. Impact in South Africa is in its infancy but will likely be substantial to local economies and ecosystems. Mitigation has proven difficult, but numerous research projects have been initiated throughout the country.

## Study species

Ambrosia beetles in the tribe Xyleborini (Coleoptera: Curculionidae: Scolytinae) are emerging as a group of globally important invasive insect pests.[1] *Euwallacea,* a genus within the Xyleborini, contains 45 species, including four lineages in the *E. fornicatus* sensu lato species complex: *E. fornicatior*, *E. perbrevis*, *E. kuroshio* and *E. fornicatus*. The taxonomy of the *E. fornicatus* species complex has been described as one of the most challenging in scolytine systematics[2]; however, the species complex has recently been resolved, with the name *E. fornicatus* now formally linked to the common name, polyphagous shot hole borer (PSHB) (Figure 1)[2-4]. Diagnostic characters separating these species have been described, and while the four clades on average differ morphologically, notable overlap in morphological characters precludes complete delimitation.[2,4] The most reliable means of identification is therefore via DNA sequence comparisons[2,4] or high-resolution melt assay of the 28S ribosomal subunit[5].

PSHB originates from Southeast Asia[3,6], and, like other ambrosia beetles, has a mutualistic association with fungi. Three fungal symbionts have been described from PSHB and infected plant hosts: *Fusarium euwallaceae*, *Graphium euwallaceae* and *Paracremonium pembeum.*[7-9] These fungal symbionts are introduced into the galleries constructed by the beetles in host trees, via mandibular mycangia – specialised internal organs evolved for fungal transport.[7-9] Most notable of these three fungi is the primary symbiont *Fusarium euwallaceae*. In addition to being the main food source for the beetle and its larvae, the fungus invades the vascular tissue of the tree, blocking the transport of water and nutrients, and resulting in Fusarium dieback.[7-9] The majority of ambrosia beetles typically infest stressed, dead or dying trees; however, species within the *E. fornicatus* species complex are amongst the very few ambrosia beetles capable of infesting healthy plants[10], and therefore present an important phytosanitary threat in their invaded range.

## History of invasion

PSHB was first detected outside its native range in 2003 in urban California, USA, reported as the tea shot hole borer.[11] In 2009, it was detected on avocado in Israel (as *Euwallacea* sp. nr *fornicatus*), but anecdotal evidence suggests that it was introduced as early as 2004.[12] It is now known to infest avocado trees and many other woody plants in urban and agricultural areas in these countries.[12,13] Even though exotic *Euwallacea* has been known from Hawaii since 1910[14], *E. fornicatus* was only confirmed among populations of *E. perbrevis* in 2020[5,15]. PSHB was also recently (April 2020) detected in a greenhouse in Italy; the pest status is currently: present, under eradication.[16]

In 2017, as part of a plant sentinel project, PSHB was found infesting plane trees (*Platanus* x *acerifolia*) in a botanical garden in the KwaZulu-Natal Province, South Africa.[17] However, a DNA sequence of PSHB was obtained from a specimen in Durban in 2012 as part of the Barcode of Life project (BOLD: ACC9773, ETKC270–13).[3] This location is ca. 50 km southeast of the 2017 collection site, indicating a much earlier introduction into the country. Since then, the beetle has been found in eight of the nine South African provinces, reaching as far as 1000 km from the first detection locality, making it the largest current outbreak of this pest globally.


*Scale bars = 1 mm*

**Figure 1:** (a) Adult and (b) larva of the polyphagous shot hole borer beetle, *Euwallacea fornicatus*.

## Factors driving invasion by PSHB in South Africa

### *Introduction and dispersal pathways*

Central to the understanding of the likelihood of establishment of species beyond their native range are the characters that enhance their dispersal, i.e. departure, transport, and arrival in non-native environments.[18] For PSHB, chances for departure are high due to their small size (females are 1.8–2.5 mm long while males are only 1.5–1.67 mm long) and cryptic colouration (dark to light brown).[19] Therefore, they are easily missed during routine inspections when hitchhiking on goods and biological material such as nursery stock, even when not colonising the material.[20] Their cryptic life cycle adds considerably to their successful departure as contaminants that colonise wood products.[21] Ambrosia beetles typically infest host trees by constructing galleries deep within the vascular tissues.[22] They are therefore easily missed in untreated wooden articles and packaging materials.[23] Potential transport vectors for PSHB include aeroplanes, trains, trucks, cars, and boats. PSHB was most likely introduced into South Africa via ships as the earliest known record for it is close to two major ports.[3] Subsequent spread throughout the country was probably aided by the movement of infested wood (e.g. firewood).

As hitchhikers, characters such as resistance to drastic temperatures, starvation, dehydration, or exposure to toxins enhance survival during the transport phase.[18,24] As there is no evidence that PSHB is particularly resistant to any of these conditions, it likely survives during transport as contaminants due to their colonisation of wooden products.[25] Wood acts as a buffer against rapid temperature fluctuations and is therefore a good insulator. Additionally, many wood types are not particularly porous, preventing chemical treatments from penetrating far enough to reach the beetle colony. Finally, moisture content is often fairly high, benefitting fungal growth and beetle survival[26], but adults of PSHB are often found in fairly dry wood, where its second nutritional symbiont *Graphium euwallaceae* thrives[8]. As contaminants of wood, individuals may be resistant to phytosanitary treatments[27] and are difficult to manage through application of a single phytosanitary measure[23]. Their small size and cryptic colouration also aid in preventing detection during inspections.[28] In South Africa, there is an increased risk of accidental introductions due to limited phytosanitary inspections – a consequence of a lack of trained personnel and funding.[29]

### *Reproduction and propagule pressure*

High propagule pressure can enhance the chances of dispersal at all three levels (departure, transport, and arrival).[30-32] Studies rearing PSHB on artificial diets have provided valuable data on beetle biology, phenology, sex ratios and temperature requirements.[25,33] Development rates for PSHB are temperature dependent, with $t_{min}$, $t_{max}$ and $t_{opt}$ determined to be 13.34 °C, 33.08 °C and 27.51 °C, respectively, with the number of degree days required for development from egg to adult estimated to be 398 ± 52.23.[33] Fungal growth is also influenced by temperature, with optimum growth of *F. euwallaceae* occurring around 27 °C (Freeman S 2020, personal communication). Fungal growth rate is likely to play a role in brood size and may explain the observation of reduced brood sizes at higher temperatures.[33]

A single foundress female lays an average of 32.4 eggs, with the first adults developing in as little as 22 days at 24 °C.[25] For PSHB, outbreeding is not required for reproduction; it has a haplodiploid mating system in which haploid males develop from unfertilised eggs.[25] If a mated female initiates a colony, both male and female offspring may be produced. Sex ratios are heavily female biased, with only ca. 7% male beetles.[25] These male offspring can mate with female offspring within the galleries, resulting in mated female progeny that can leave the natal gallery and start new colonies. However, a virgin female can construct a gallery by laying haploid eggs that will result in all-male offspring.[25] She can then mate with her male offspring to produce diploid eggs.[25] The result is that a single female beetle, whether mated or not, can start a new infestation with no indication that the population will suffer from inbreeding.[3,34] For PSHB, successful establishment is therefore possible even with low propagule pressure.[4]

Colonies contain multiple overlapping generations and a colony from a single foundress can produce up to 57 reproductive female adults within 6–7 weeks.[25] Infested wood generally contains multiple colonies from numerous foundress females and population sizes can be tremendously large and self-sustaining over relatively long periods of time. For example, a single piece of infested wood weighing ca. 500 g collected in South Africa produced 44 mature females and five males after drying. These adult females are able to produce ca. 1425 offspring (ca. 32.4 eggs per female) within 22 days at 24 °C if all are able to successfully construct galleries.[25] It is clear that propagule pressure for PSHB is high, enhancing its chances of dispersal and establishment.[35] It has also been shown that adult beetles and larvae are able to survive in cut wood, with females continuing to lay viable eggs.[19] Studies from California have shown that beetles can continue to emerge for several months after felling of infested trees.[19]

### *Host range*

PSHB is known to have a particularly broad host range, in which both the fungus and beetle are able to establish.[36-38] This increases the chances of a suitable host being encountered, thus increasing the chances of successful establishment. To obtain the PSHB host range for South Africa, observations made by researchers during field investigations and reports made by the public to the Forestry and Agricultural Biotechnology Institute, University of Pretoria, and Stellenbosch University were collated. Only taxa from which the primary fungal symbiont, *F. euwallaceae*, or a beetle individual were successfully isolated and identified using DNA sequence fingerprinting were included. Reproductive hosts include only those hosts that have been shown to be suitable for PSHB reproduction at least once (i.e. those with galleries containing eggs, larvae or

multiple adults). Non-reproductive hosts include those from which either PSHB or *F. euwallaceae* was isolated, but no evidence of PSHB reproduction was found. In South Africa, signs of PSHB infestations have been confirmed from 130 tree species of which 48 are suitable for reproduction (i.e. reproductive hosts) and 82 can sustain fungal growth (i.e. non-reproductive hosts) (Table 1). This wide fungal host range ensures availability of suitable hosts near the introduction point. When beetle populations are very high, the insects appear to assess newly encountered hosts almost indiscriminately for fungal establishment[13,38], explaining the high numbers of non-reproductive hosts. Such 'attacks' may also lead to severe decline in tree health due to Fusarium wilt[13]. *Fusarium euwallaceae* may also help PSHB to overcome tree defences and enhance successful colonisation of a host, even those that are only marginally suitable for beetle development.[39]

### *Distribution of PSHB in South Africa*

As with the determination of the PSHB host range in South Africa, the PSHB distribution was mapped by collating confirmed (by DNA sequence fingerprinting) reports based on field observations made by researchers and reports made by the public. In South Africa, PSHB is widespread, and has a continuous distribution in both KwaZulu-Natal (between Durban and Richards Bay) and the southern Cape (between George and Plettenberg Bay) (Figure 2). This indicates a relatively long history in these areas where initial populations expanded from the points of introduction. Unlike males, females can fly, but most disperse only short distances (ca. 30 m from point of release[40]). This may seem unremarkable, but factors like wind can have tremendous influence on the number of dispersing individuals, the direction of dispersal and the dispersal distance.[40] In Somerset West (Western Cape Province), *E. fornicatus* dispersed at least 3 km from the putative point of introduction in only 2 months, against prevailing winds. Therefore, when suitable hosts are present, PSHB can expand its range rapidly without direct human assistance. However, distances between closest confirmed PSHB populations in South Africa can be as far as 500 km (Figure 2). This can only be achieved when dispersal is aided by humans and human-mediated dispersal remains the most substantial form of long-distance transport and establishment of new colonies.

PSHB has a wide range of climatic conditions in which it can persist, increasing its chances of finding sites open for establishment. It is native to equatorial climates and has successfully established in temperate (California and Israel) and tropical (Hawaii) regions. Currently occupied climate zones, as defined by Köppen-Geiger and updated by Peel et al.[41], include Af, Am, Aw, BSk, Csa, Csb, Cfa, Cfb and Cwa. Within South Africa, PSHB has established in at least six climatic zones: BSh, BSk, Cfa, Cfb, Csb and Cwb. In addition, Csa and Cwa are also present in South Africa, making much of the country climatically suitable for PSHB. While the minimum temperature threshold for PSHB reproduction has been determined to be 13.34 °C, beetles are able to persist at lower temperatures for extended periods.[33] A cold tolerance study found significant mortality rates among colonies exposed to −5 °C and −1 °C; however, no mortality was observed for colonies exposed to 0 °C, 1 °C or 5 °C.[25] In addition, PSHB galleries are often located deep within wood, where they are protected from extreme temperatures. Finally, microclimates exist in urban and agricultural settings, which may provide a conducive climate in an otherwise unsuitable area.

In South Africa, PSHB is currently mostly confined to urban environments, likely due to greater invasibility and anthropogenically mediated dispersal. Also, urban areas are far less biologically diverse than natural environments and offer less of a biological buffer.[42] Typically, ambrosia beetles favour trees that are stressed in their natural environments.[43] Even in novel environments this is often the case. Anthropogenic activities such as urbanisation, gardening, trampling, and pruning can cause tree stress.[44] As a result, urban areas have high invasibility.[42,45] Natural disturbances such as drought, storms, fire, and other pests can also produce added stress to trees in both urban and natural settings.[46] Combined, these factors promote the spread and status of PSHB, but none has yet been comprehensively evaluated. Although PSHB is unlike many other ambrosia beetle species, in that it can easily infest healthy hosts[10,47], if tree stress is still a contributing factor to epidemiology, a natural or anthropogenic disturbance in future could have catastrophic consequences.

## Impact of PSHB

The first indication of PSHB in invaded habitats is benign. It tends to infest healthy hosts where 0.85-mm diameter holes penetrate through the bark directly into the sapwood (Figure 3a). Lesions may be visible on the outside of the tree (Figure 3b), but more commonly a brown to pinkish stain, resulting from colonisation by *F. euwallaceae*, can be seen spreading from the gallery into the vascular tissues below the bark (Figure 3c). Different hosts respond to infestation in different ways, thus complicating initial diagnosis. External signs of infestations include wet patches and gum exudates (tree resin flows), sugar fountains, and accumulating frass (mixture of sawdust and excreta) or noodles (compacted frass) (Figures 3d–g). Successful colonisation of reproductive hosts often results in wilting and death of the infested branch or the entire tree (Figure 3h).[12,36] By this stage, beetle populations are well established. Detection in natural systems is very challenging due to a greater tree diversity, fewer people, and larger areas to survey.

The most notable agricultural impact of PSHB invasion has been on the avocado industries of Israel and California, USA. In Israel, it was detected in orchards in 2009, and by 2016 it had spread to nearly all avocado-growing areas in the country.[48] Symptoms such as stem and branch breakages, wilting, dieback and death are common.[48] Initial studies from Israel and California reported a high risk of branch dieback and mortality, raising concern about the potential of PSHB to severely impact avocado production worldwide.[7,12,49] More recent observations, however, suggest impacts have not been as severe as anticipated, and that sanitation practices, including the removal of infested branches and trees, may offer a cost-effective management approach.[38,48] Coleman et al.[50] emphasise though, that their survey data represent only one time point and that further surveys are needed to confirm potential impacts to production. Caution should be taken in extrapolating this to the situation in South Africa due to differences in climates, cultivation practices and the relative importance of different cultivars for production. South Africa currently has around 17 500 ha avocado orchards, producing around 170 000 tons (or around ZAR700 million worth) of avocado annually. However, to date, no PSHB infestations have been found in commercial avocado orchards (only in backyard avocado trees). The only agricultural crops thus far affected by PSHB infestation in the country appear to be pecan and macadamia, but the effect on these crops appears limited, most likely as they are not reproductive hosts.

PSHB has also been found associated with numerous other agricultural crops such as cherry, apple, citrus, peach, guava, olive, grape vine and prune (Table 1). Whether any of these can act as reproductive hosts is still unclear as all reports are from urban settings. Even so, when grown near reproductive hosts, there is a strong possibility of infection by *F. euwallaceae*. Therefore, current research in South Africa is aimed at establishing whether these crops are susceptible to Fusarium wilt. Besides agriculturally important tree species, PSHB have been observed on *Acacia mearnsii*, a commercially grown forestry species in South Africa, but to date, no trees in commercial plantations have been reported to be infested. Based on observations on *Acacia* spp. from plantations in southeast Asia, there are concerns surrounding the threat posed to the South African forestry industry.[51]

In its global invaded range, the most notable impacts of PSHB are on urban trees, with hundreds of different species susceptible to beetle infestation and Fusarium wilt[13,38,48] (Table 1). In the USA, nearly 300 tree species have been reported to be susceptible to infestation by PSHB, with 55 of these suitable for reproduction.[38] PSHB has caused injury and/or mortality to thousands of ornamental trees in urban forests in southern California.[38] In South Africa, visual surveys undertaken in invaded urban areas of Johannesburg (Gauteng Province), Knysna, George and Somerset West (Western Cape Province) showed that nearly all *Quercus robur* (English oak), *Acer negundo* (box elder) and other maples (*Acer* spp.) will die when infested by PSHB (Z.W.d.B., T.P., F.R., unpublished data). Currently about 50% of London plane trees (*Platanus* x *acerifolia*) infested with PSHB in these areas are in decline and likely to die.

**Table 1:** List of confirmed host plants of the polyphagous shot hole borer beetle in South Africa. The list includes only those taxa from which its primary fungal symbiont, *Fusarium euwallaceae*, or a beetle individual was successfully isolated and its identity confirmed using DNA sequence fingerprinting. Reproductive hosts include only those hosts that have been shown to be suitable for PSHB reproduction at least once. Non-reproductive hosts include those from which either PSHB or *F. euwallaceae* was isolated, but no evidence of PSHB reproduction has been verified.

| Species | Family | Common name | Indigenous or exotic in South Africa |
|---|---|---|---|
| **Reproductive hosts** | | | |
| *Acacia longifolia* | Fabaceae | Long-leafed wattle | Exotic |
| *Acacia mearnsii* | Fabaceae | Black wattle | Exotic |
| *Acacia melanoxylon* | Fabaceae | Blackwood | Exotic |
| *Acer buergerianum* | Aceraceae | Trident (Chinese) maple | Exotic |
| *Acer negundo* | Aceraceae | Boxelder | Exotic |
| *Acer palmatum* | Aceraceae | Japanese maple | Exotic |
| *Acer saccharinum* | Aceraceae | Silver maple | Exotic |
| *Afzelia quanzensis* | Fabaceae | Pod mahogany | Indigenous |
| *Anisodontea scabrosa* | Malvaceae | Rough-leaf African mallow | Indigenous |
| *Bauhinia galpinii* | Fabaceae | Pride of De Kaap | Indigenous |
| *Bauhinia variegata* | Fabaceae | Orchid tree | Exotic |
| *Brachychiton discolor* | Malvaceae | Pink flame tree | Exotic |
| *Brachylaena discolor* | Asteraceae | Coast silver oak | Indigenous |
| *Calpurnia aurea* | Fabaceae | Wild laburnum | Indigenous |
| *Casuarina cunninghamiana* | Casuarinaceae | Beefwood | Exotic |
| *Combretum erythrophyllum* | Combretaceae | River bushwillow | Indigenous |
| *Combretum krausii* | Combretaceae | Forest bushwillow | Indigenous |
| *Diospyros glabra* | Ebenaceae | Cape star-apple | Indigenous |
| *Erythrina caffra* | Fabaceae | Coral tree | Indigenous |
| *Gleditsia triacanthos* | Fabaceae | Honey locust | Exotic |
| *Kiggelaria africana* | Achariaceae | Wild peach | Indigenous |
| *Liquidambar styraciflua* | Altingiaceae | Sweetgum | Exotic |
| *Magnolia grandiflora* | Magnoliaceae | Southern magnolia | Exotic |
| *Persea americana* | Lauraceae | Avocado | Exotic |
| *Photinia x fraseri* | Rosaceae | Christmas berry | Exotic |
| *Platanus x acerifolia* | Platanaceae | London plane | Exotic |
| *Podalyria calyptrata* | Fabaceae | Water blossom pea | Indigenous |
| *Populus alba* | Salicaceae | White poplar | Exotic |
| *Populus x canescens* | Salicaceae | Grey poplar | Exotic |
| *Prunus cerasifera* | Salicaceae | Lombardy poplar | Exotic |
| *Populus simonii* | Salicaceae | Chinese cottonwood | Exotic |
| *Prunus nigra* | Rosaceae | Black plum | Exotic |
| *Psoralea aphylla* | Fabaceae | Leafless fountain bush | Indigenous |
| *Psoralea pinata* | Fabaceae | Fountain bush | Indigenous |
| *Quercus palustris* | Fagaceae | Pin oak | Exotic |
| *Quercus robur* | Fagaceae | English oak | Exotic |
| *Quercus suber* | Fagaceae | Cork oak | Exotic |
| *Ricinus communis* | Euphorbiaceae | Castor bean | Exotic |
| *Salix alba* | Salicaceae | White willow | Exotic |
| *Salix mucronata* | Salicaceae | Cape willow | Indigenous |
| *Sparrmannia africana* | Malvaceae | African hemp | Indigenous |
| *Trema orientalis* | Cannabaceae | Pigeon wood | Indigenous |
| *Trichilia emetica* | Meliaceae | Natal mahogany | Indigenous |
| *Ulmus parvifolia* | Ulmaceae | Chinese elm | Exotic |
| *Vepris lanceolata* | Rutaceae | White ironwood | Indigenous |
| *Viburnum odoratissimum* | Adoxaceae | Sweet viburnum | Exotic |
| *Virgilia oroboides* | Fabaceae | Keurboom | Indigenous |
| *Wisteria sinensis* | Fabaceae | Chinese wisteria | Exotic |
| **Non-reproductive hosts** | | | |
| *Adansonia digitata* | Malvaceae | Baobab | Indigenous |
| *Afrocarpus falcatus* | Podocarpaceae | Outeniqua yellowwood | Indigenous |
| *Albizia adianthifolia* | Fabaceae | Flat crown | Indigenous |
| *Bauhinia purpurea* | Fabaceae | Butterfly orchid tree | Exotic |
| *Betula pendula* | Betulaceae | Silver birch | Exotic |
| *Bougainvillea* | Nyctaginaceae | Bougainvillea | Exotic |
| *Buddleja saligna* | Scrophulariaceae | False olive | Indigenous |
| *Calodendrum capense* | Rutaceae | Cape chestnut | Indigenous |
| *Camellia japonica* | Theaceae | Common camellia | Exotic |
| *Carya illinoinensis* | Juglandaceae | Pecan | Exotic |
| *Ceiba pentandra* | Malvaceae | Kapok | Exotic |
| *Celtis africana* | Cannabaceae | White stinkwood | Indigenous |
| *Cinnamomum camphora* | Lauraceae | Camphor tree | Exotic |
| *Citrus limon* | Rutaceae | Lemon | Exotic |
| *Citrus sinensis* | Rutaceae | Orange | Exotic |
| *Commiphora harveyi* | Burseraceae | Copper-stem corkwood | Indigenous |

**...Table 1 continues on next page**

**Table 1 continued...**

| Species | Family | Common name | Indigenous or exotic in South Africa |
|---|---|---|---|
| **Non-reproductive hosts** | | | |
| *Cordia caffra* | Boraginaceae | Septee tree | Indigenous |
| *Cussonia spicata* | Araliaceae | Cabbage tree | Indigenous |
| *Diospyros dichrophylla* | Ebenaceae | Star apple | Indigenous |
| *Diospyros kaki* | Ebenaceae | Persimmon | Exotic |
| *Diospyros lycioides* | Ebenaceae | Monkey plum | Indigenous |
| *Diospyros whyteana* | Ebenaceae | Bladdernut | Indigenous |
| *Dombeya rotundifolia* | Malvaceae | Wild pear | Indigenous |
| *Dovyalis caffra* | Salicaceae | Kei apple | Indigenous |
| *Ekebergia capensis* | Meliaceae | Cape ash | Indigenous |
| *Eriobotrya japonica* | Rosaceae | Loquat | Exotic |
| *Erythrina livingstoniana* | Fabaceae | Aloe coral tree | Exotic |
| *Erythrina lysistemon* | Fabaceae | Common coral | Indigenous |
| *Eucalyptus camaldulensis* | Myrtaceae | River red gum | Exotic |
| *Ficus carica* | Moraceae | Common fig | Exotic |
| *Ficus natalensis* | Moraceae | Natal fig | Indigenous |
| *Ficus sur* | Moraceae | Cape fig | Indigenous |
| *Fraxinus americana* | Oleaceae | American ash | Exotic |
| *Fraxinus excelsior* | Oleaceae | European ash | Exotic |
| *Grewia occidentalis* | Malvaceae | Cross berry | Indigenous |
| *Gymnosporia buxifolia* | Celestraceae | Common spikethorn | Indigenous |
| *Hakea salicifolia* | Proteacea | Willow-leaved hakea | Exotic |
| *Halleria lucida* | Stilbaceae | Tree fuchsia | Indigenous |
| *Harpephyllum caffrum* | Anacardiaceae | Wild plum | Indigenous |
| *Hibiscus rosa-sinensis* | Malvaceae | Chinese hibiscus | Exotic |
| *Ilex mitis* | Aquifoliaceae | Cape holly | Indigenous |
| *Jacaranda mimosifolia* | Bignoniaceae | Jacaranda | Exotic |
| *Leonotis leonurus* | Lamicaeae | Wild tobacco | Indigenous |
| *Macadamia* | Proteacea | Macadamia | Exotic |
| *Malus domestica* | Rosaceae | Apple | Exotic |
| *Melia azedarach* | Meliaceae | Syringa | Exotic |
| *Metasequoia glyptostroboides* | Cupressaceae | Dawn redwood | Exotic |
| *Morus nigra* | Moraceae | Black mulberry | Exotic |
| *Nuxia floribunda* | Stilbaceae | Forest elder | Indigenous |
| *Olea europaea* subsp. *africana* | Oleaceae | Wild olive | Indigenous |
| *Olea europaea* subsp. *europaea* | Oleaceae | Cultivated olive | Exotic |
| *Olinia ventosa* | Penaeaceae | Hard pear | Indigenous |
| *Osteospermum moniliferum* | Asteraceae | Bietou | Indigenous |
| *Platanus occidentalis* | Platanaceae | American plane | Exotic |
| *Platanus racemosa* | Platanaceae | Californian plane | Exotic |
| *Plumeria rubra* | Apocynaceae | Frangipani | Exotic |
| *Podocarpus henkelii* | Podocarpaceae | Henkel's yellowwood | Indigenous |
| *Protea mundii* | Proteacea | Forest sugar bush | Indigenous |
| *Prunus africana* | Rosaceae | African cherry | Indigenous |
| *Prunus avium* | Rosaceae | Sweet cherry | Exotic |
| *Prunus cerasifera* | Rosaceae | Cherry plum | Exotic |
| *Prunus persica* | Rosaceae | Peach | Exotic |
| *Psidium guajava* | Myrtaceae | Guava | Exotic |
| *Quercus rugosa* | Fagaceae | Net leaf oak | Exotic |
| *Rapanea melanophloeos* | Primulaceae | Cape beech | Indigenous |
| *Robinia psuedoacacia* | Fabaceae | Black locust | Exotic |
| *Salix babylonica* | Salicaceae | Weeping willow | Exotic |
| *Schinus molle* | Anacardiaceae | Pepper tree | Exotic |
| *Schotia brachypetela* | Fabaceae | Weeping boer-been | Indigenous |
| *Searsia chirindensis* | Anacardiaceae | Red currant | Indigenous |
| *Searsia lansea* | Anacardiaceae | Karree | Indigenous |
| *Senegalia burkei* | Fabaceae | Black monkey thorn | Indigenous |
| *Senegalia galpinii* | Fabaceae | Monkey thorn | Indigenous |
| *Solanum mauritianum* | Solanaceae | Bugweed | Exotic |
| *Syzygium cordatum* | Myrtaceae | Waterberry | Indigenous |
| *Taxodium distichum* | Cupressaceae | Bald cypress | Exotic |
| *Ulmus procera* | Ulmaceae | English elm | Exotic |
| *Vachellia karroo* | Fabaceae | Sweet thorn | Indigenous |
| *Vachellia sieberiana* var. *woodii* | Fabaceae | Paperbark thorn | Indigenous |
| *Virgilia divaricata* | Fabaceae | Keurboom | Indigenous |
| *Vitis vinifera* | Vitaceae | Grape | Exotic |

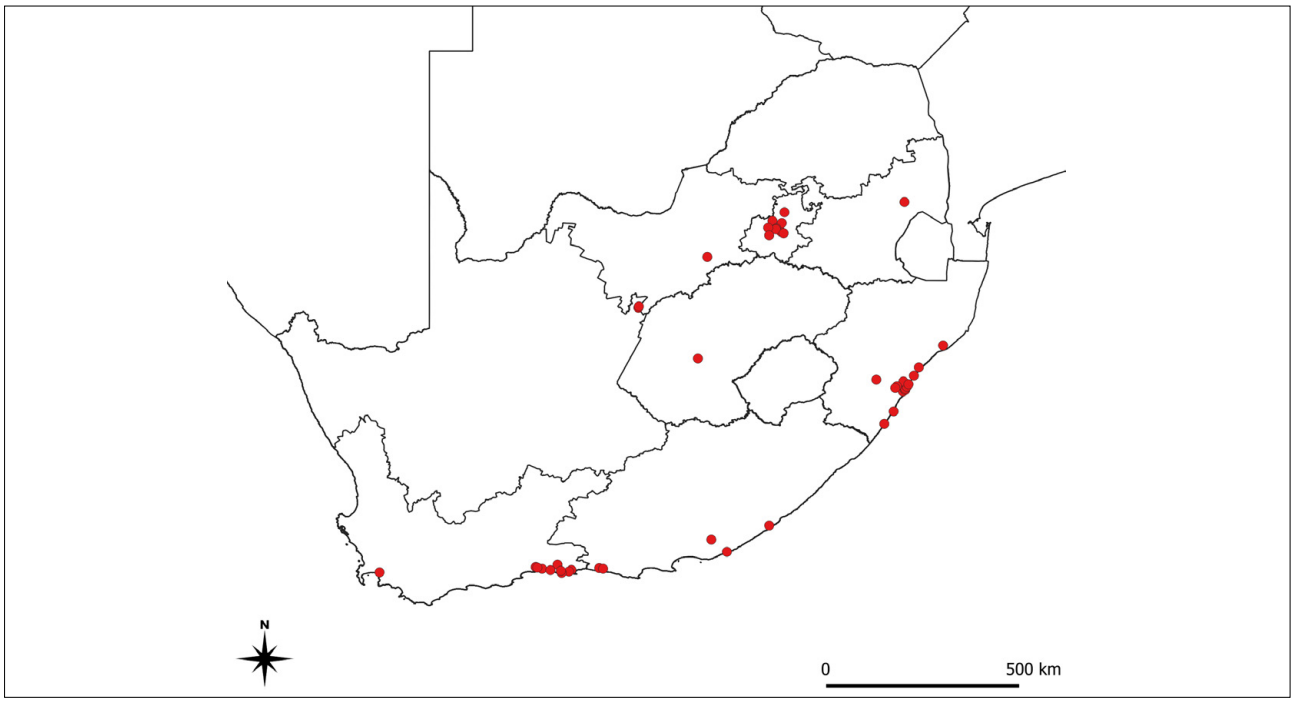



**Figure 2:** Distribution of the polyphagous shot hole borer beetle (PSHB; *Euwallacea fornicatus*) in South Africa. Localities depicted are only those from which PSHB or its symbiotic fungus *Fusarium euwallaceae* were collected and their identities verified using DNA sequencing.

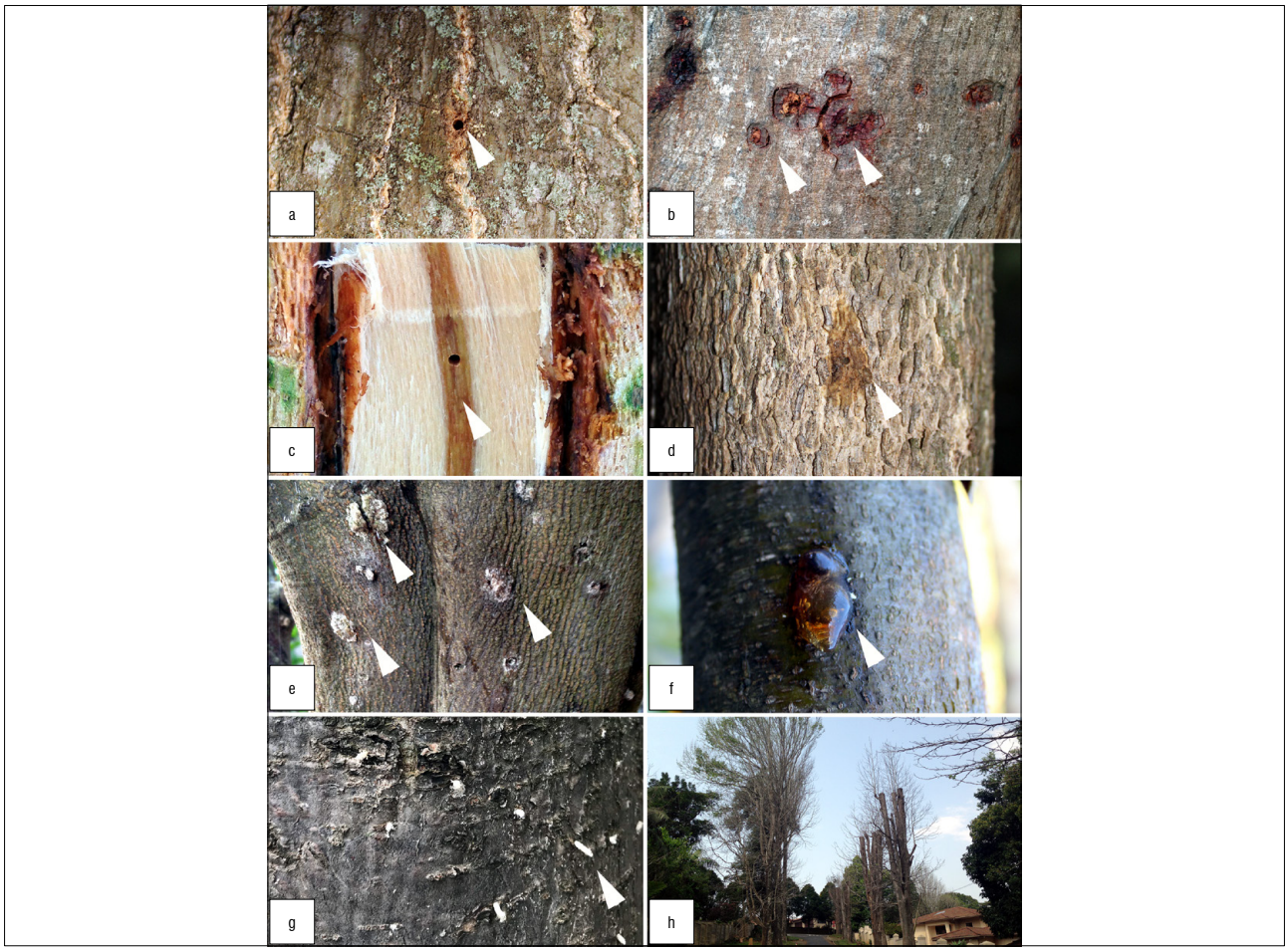



**Figure 3:** Signs of infestations of polyphagous shot hole borer (PSHB) beetle, *Euwallacea fornicatus,* on various hosts: (a) ca. 0.85-mm diameter entrance hole of PSHB on the bark of *Erythrina caffra*. (b) Lesions on the bark of *Acacia mearnsii* around entrance holes of PSHB. (c) Colonisation of vascular tissues by *Fusarium euwallaceae* (staining) surrounding the entrance hole of PSHB under the bark of *Platanus* x *acerifolia*. (d) Wet patch on the bark of *Nuxia floribunda* following infestation by PSHB. (e) Sugar fountain (dried sugary sap) on the bark of *Grewia occidentalis* following infestation by PSHB. (f) Gum exudates (tree resin flow) on the bark of *Podalyria calyptrata* following infestation by PSHB. (g) Noodles (compacted frass) extruded from the galleries of PSHB following infestation of *Harpephyllum caffrum*. (h) *Liquidambar styraciflua* individuals showing dieback of branches and tree death following heavy infestations by PSHB.

Unfortunately, data on the species and numbers of trees planted in South African cities are lacking, and, for many of the reproductive host tree species planted in urban environments, no data are available on the percentage of trees that will die. However, a recent system dynamics model to estimate the potential economic impact of PSHB in South African urban environments could be substantial, including impacts on stormwater run-off, losses to ecosystem services and carbon dioxide sequestration, and reduced shade contributing to remediation of the urban heat island effect and pollution reduction.[52] There is increasing evidence linking widespread death of trees in urban environments following pest and pathogen invasions, and increased human mortality related to cardiovascular and respiratory tract illness.[53] This may be particularly relevant in large cities with an urban/industrial interface where trees play an important role in pollution remediation. As trees succumb, falling trees and branches may impact human safety directly by causing traffic accidents and damaging vehicles and infrastructure such as buildings, fences, and power lines.

Negative economic impacts are incurred with the removal of infested trees and their replacement. However, tree removal has additional negative effects, such as a possible reduction in property values. In South Africa this has already resulted in at least one court case where PSHB infestations were not reported to potential buyers. The cost of tree removal in Somerset West amounted to just over ZAR7000/tree in 2019. With more than 600 trees earmarked for removal in this town alone, the City of Cape Town requested a stipend of ZAR3 million in 2020 to monitor and remove infested trees (Ramabulana P 2019, personal communication). Services provided by these infested trees (e.g. carbon sequestering, providing shade, aesthetics, and emotional health) are also severely negatively affected by their removal. The replacement cost for trees that have already been removed in Somerset West is estimated to be between ZAR4.5 million and ZAR5.8 million, depending on tree species and labour costs (E.v.R., unpublished data). Therefore, by far the largest proportion of the social cost due to PSHB invasion in South Africa will be incurred in urban areas.[52]

Most worrying perhaps is the ability of PSHB to infest and kill native trees in its invaded ranges. Of the reproductive host species recorded in South Africa, 19 are indigenous (Table 1), and severe ecological consequences may result after invasion of natural environments, as was seen with the destruction of endemic willow forests in the Tijuana River Valley.[54] Approximately 120 000 (30%) native willows (*Salix* spp.) were killed following infestation by *E. kuroshio*, a close relative of PSHB.[55] Loss of shade resulted in increased river temperatures, which negatively impacted aquatic organisms. Recovery from the loss of large established trees will take time and encroachment by invasive alien plants has prevented full recovery of some stands.[56] The invasion of PSHB into South Africa's natural forests is still in its infancy, and there is no empirical data yet on the numbers of trees expected to succumb. Long-term monitoring plots have, however, been established in natural forests in Knysna and George, with some preliminary data available from ongoing surveys in these locations. In the ca. 3000 km$^2$ surveyed, 2195 trees were assessed for PSHB presence. Of these, 217 individuals (ca. 10%) were infested and 191 (8.7%) showed signs of beetle reproduction, which may lead to dieback and death (E.v.R., G.T., unpublished data). In South Africa, the loss of even a single species from an ecosystem can be catastrophic. For example, *Virgilia oroboides* is a native reproductive host for PSHB. It is an important pioneer forest species in the southern Cape that protects forests from severe climatic fluctuations and fire and houses a large number of native organisms.[57,58] Elimination of this single species could have irreversible consequences for native forest integrity.

## Mitigation

Experiences from California, Israel, and South Africa indicate that the management of the PSHB is particularly challenging. To summarise, miniscule size presents detection challenges, and the haplodiploid and sib-mating system means the introduction of a single female can lead to the establishment of a new infestation. As the beetle feeds on symbiotic fungi, it has low tree host specificity. It is able to persist in many climatic zones and climate may not be critical for its establishment. The reproductive cycle is fast under optimum conditions, and if temperatures fall outside the suitable range, larvae and beetles can survive for some time and continue development should temperatures become suitable again. Spread is difficult to manage due to the ease of movement of infested wood and green waste. Despite the challenges, there are several mitigation strategies that may aid in reducing the impact of PSHB.

Biological control is generally difficult for ambrosia beetles owing to their cryptic habits and inbreeding success.[48] There is no biocontrol agent available against PSHB. However, Californian researchers have been investigating PSHB in its natural range to identify potential natural enemies (Stouthamer R 2021, personal communication). In South Africa, investigations into biocontrol options are being initiated.

Control of low levels of PSHB infestation may be achievable by direct injection of insecticides and fungicides. This not only maximises pesticide impact but reduces environmental contamination. Control of PSHB in more heavily infested trees seems unattainable even with the direct application of chemicals.[59] Chemicals that show the most promise include combinations of insecticide and fungicide treatments, like emamectin benzoate or bifenthrin combined with propiconazole.[59-61] These controls are currently being investigated in South Africa. A report by the California Forest Pest Council[62], however, highlights that chemical treatment of infested trees in Orange County currently requires repeated applications (up to three times per year) using a combination of insecticides and fungicides and various application methods. The cost of chemical treatment is therefore expected to be high, and even if proven effective, their application in natural settings will not be feasible. Additionally, the application of commercial chemicals for prolonged periods could result in phytotoxicity.[63-65] In South Africa, no insecticides or fungicides that have been evaluated with scientific rigour for efficacy in PSHB management have been registered for use.

As there is no effective means to control pre-existing beetle infestations within a tree, the best alternative is to reduce the propagule number by physical removal of highly infested reproductive host trees. This reduces propagule pressure and potential of spread to new localities. This strategy has been used to reduce PSHB impact in infested avocado orchards.[38,48] Infested material needs to be properly treated to reduce beetle survival. Currently, best protocols dictate chipping of the wood to pieces smaller than 5 cm followed by solarisation under plastic sheeting for 4–6 weeks at 55 °C.[19,66] Movement of infested wood and wood chips should be avoided.[32,67] Minimising the spread of invasive reproductive host plant species is crucial, as these may create pathways for autonomous spread.[68]

## Conclusions

PSHB has proven to be one of the most difficult tree pests to manage. This is largely due to its ability to colonise a massive variety of tree species and its ease of anthropogenically mediated long-distance dispersal. In the absence of a rapidly implemented nationally coordinated strategy, and without strong stakeholder and public engagement, PSHB has easily spread throughout the country.[69] Thus far, it has invaded numerous major cities, *and*, because South Africa is climatically suitable, PSHB is expected to rapidly spread to new areas. Impacts of the invasion are currently only in their infancy and major social and ecological costs are expected in the near future.[52] Continued monitoring of the spread of the beetle is essential if we are to mitigate its impact. Mitigation will prove difficult, but is of the utmost importance. Even halting the movement of infested wood and wood products will be problematic. Despite the informal urban firewood trade being a widespread and important energy and income security source for many poor urban residents, its prevention is seen as the most effective way to curb long-distance spread. One benefit of PSHB invasion is an increase in public awareness of invasive organisms and their potential impact, at least in heavily affected areas. This increase in public awareness has also provided an opportunity in terms of public education and, combined with the creation of a multidisciplinary and multi-institutional Polyphagous Shot Hole Borer Research Network in South Africa, will help considerably in continued efforts in research, monitoring and control of invasive species.

## Acknowledgements

We thank the Natural Resource Management Programme of the South African Department of Forestry, Fisheries and the Environment (DFFE) for funding, noting that this publication does not necessarily represent the views or opinions of DFFE or its employees. We acknowledge the Tree Protection Cooperative Programme (TPCP) and the Forestry and Agricultural Biotechnology Institute (FABI) of the University of Pretoria for funding and infrastructure support. Comments from anonymous reviewers helped to improve our manuscript.

## Competing interests

We declare that there are no competing interests.

## Authors' contributions

E.v.R.: Conceptualisation; methodology; data collection; validation; data curation; writing – the initial draft; writing – revisions. T.P.: Data collection; validation; data curation; writing – revisions. W.d.B.: Conceptualisation; methodology; data collection; validation; data curation; student supervision; project management; and funding acquisition. G.T.: Data collection; validation; writing – revisions. S.F.: Data collection; validation; writing – revisions. W.J.N.: Data collection; validation; writing – revisions. S.M. Data collection; validation; writing – revisions. M.H.: Conceptualisation; methodology; data collection; validation; data curation; writing – revisions; project management; and funding acquisition. A.G.: Methodology; data collection; validation; data curation; writing – revisions. F.R.: Conceptualisation; methodology; data collection; validation; data curation; writing – revisions; student supervision; project leadership; project management; and funding acquisition.

## References

1. Hulcr J, Stelinski LL. The ambrosia symbiosis: From evolutionary ecology to practical management. Annu Rev Entomol. 2017;62:285–303. https://doi.org/10.1146/annurev-ento-031616-035105
2. Smith SM, Gomez DF, Beaver RA, Hulcr J, Cognato AI. Reassessment of the species in the *Euwallacea fornicatus* (Coleoptera: Curculionidae: Scolytinae) complex after the rediscovery of the "lost" type specimen. Insects. 2019;10(9):1–11. https://doi.org/10.3390/insects10090261
3. Stouthamer R, Rugman-Jones P, Thu PQ, Eskalen A, Thibault T, Hulcr J, et al. Tracing the origin of a cryptic invader: Phylogeography of the *Euwallacea fornicatus* (Coleoptera: Curculionidae: Scolytinae) species complex. Agric For Entomol. 2017;19(4):366–375. https://doi.org/10.1111/afe.12215
4. Gomez DF, Skelton J, Steininger MS, Stouthamer R, Rugman-Jones P, Sittichaya W, et al. Species delineation within the *Euwallacea fornicatus* (Coleoptera: Curculionidae) complex revealed by morphometric and phylogenetic analyses. Insect Syst Divers. 2018;2(6):1–11. https://doi.org/10.1093/isd/ixy018
5. Rugman-Jones PF, Au M, Ebrahimi V, Eskalen A, Gillett CPDT, Honsberger D, et al. One becomes two: Second species of the *Euwallacea fornicatus* (Coleoptera: Curculionidae: Scolytinae) species complex is established on two Hawaiian Islands. PeerJ. 2020;8, e9987. https://doi.org/10.7717/peerj.9987
6. Storer CG, Breinholt JW, Hulcr J. *Wallacellus* is *Euwallacea*: Molecular phylogenetics settles generic relationships (Coleoptera: Curculionidae: Scolytinae: Xyleborini). Zootaxa. 2015;3974(3):391–400. https://doi.org/10.11646/zootaxa.3974.3.6
7. Freeman S, Sharon M, Maymon M, Mendel Z, Protasov A, Aoki T, et al. *Fusarium euwallaceae* sp. nov. – a symbiotic fungus of *Euwallacea* sp., an invasive ambrosia beetle in Israel and California. Mycologia. 2013;105(6):1595–1606. https://doi.org/10.3852/13-066
8. Freeman S, Sharon M, Dori-Bachash M, Maymon M, Belausov E, Maoz Y, et al. Symbiotic association of three fungal species throughout the life cycle of the ambrosia beetle *Euwallacea* nr. *fornicatus*. Symbiosis. 2016;68(1–3):115–128. https://doi.org/10.1007/s13199-015-0356-9
9. Lynch SC, Twizeyimana M, Mayorquin JS, Wang DH, Na F, Kayim M, et al. Identification, pathogenicity and abundance of *Paracremonium pembeum* sp. nov. and *Graphium euwallaceae* sp. nov. – two newly discovered mycangial associates of the polyphagous shot hole borer (*Euwallacea* sp.) in California. Mycologia. 2016;108(2):313–329. https://doi.org/10.3852/15-063
10. Gomez DF, Lin W, Gao L, Li Y. New host plant records for the *Euwallacea fornicatus* (Eichhoff) species complex (Coleoptera: Curculionidae: Scolytinae) across its natural and introduced distribution. J Asia Pac Entomol. 2019;22(1):338–340. https://doi.org/10.1016/j.aspen.2019.01.013
11. Rabaglia RJ, Dole SA, Cognato AI. Review of American *Xyleborina* (Coleoptera: Curculionidae: Scolytinae) occurring north of Mexico, with an illustrated key. Ann Entomol Soc Am. 2006;99(6):1034–1056. https://doi.org/10.1603/0013-8746(2006)99[1034:ROAXCC]2.0.CO;2
12. Mendel Z, Protasov A, Sharon M, Zveibil A. An Asian ambrosia beetle *Euwallacea fornicatus* and its novel symbiotic fungus *Fusarium* sp. pose a serious threat to the Israeli avocado industry. Phytoparasitica. 2012;40:235–238. https://doi.org/10.1007/s12600-012-0223-7
13. Eskalen A, Stouthamer R, Lynch SC, Rugman-Jones PF, Twizeyimana M, Gonzalez A, et al. Host range of Fusarium dieback and its ambrosia beetle (Coleoptera: Scolytinae) vector in southern California. Plant Dis. 2013;97(7):938–951. https://doi.org/10.1094/pdis-11-12-1026-re
14. Swezey OH. Some injurious insects in Samoa which do not occur in the Hawaiian Islands. Proc Hawaiian Entomol Soc. 1941;11(1):127–130. http://hdl.handle.net/10125/16047
15. Mitchell A, Maddox C. Bark beetles (Coleoptera: Curculionidae: Scolytinae) of importance to the Australian macadamia industry: An integrative taxonomic approach to species diagnostics. Aust J Entomol. 2010;49(2):104–113. https://doi.org/10.1111/j.1440-6055.2010.00746.x
16. EPPO. First report of *Euwallacea fornicatus* in Italy [document on the Internet]. c2020 [cited 2021 Feb 25]. Available from: https://gd.eppo.int/reporting/article-6772
17. Paap T, De Beer ZW, Migliorini D, Nel WJ, Wingfield MJ. The polyphagous shot hole borer (PSHB) and its fungal symbiont *Fusarium euwallaceae*: A new invasion in South Africa. Australas Plant Pathol. 2018;47(2):231–237. https://doi.org/10.1007/s13313-018-0545-0
18. Gippet JM, Liebhold AM, Fenn-Moltu G, Bertelsmeier C. Human-mediated dispersal in insects. Curr Opin Insect Sci. 2019;35:96–102. https://doi.org/10.1016/j.cois.2019.07.005
19. Jones ME, Paine TD. Effect of chipping and solarization on emergence and boring activity of a recently introduced ambrosia beetle (*Euwallacea* sp., Coleoptera: Curculionidae: Scolytinae) in southern California. J Econ Entomol. 2015;108(4):1852–1859. https://doi.org/10.1093/jee/tov169
20. Liebhold AM, Brockerhoff EG, Garrett LJ, Parke JL, Britton KO. Live plant imports: The major pathway for forest insect and pathogen invasions of the US. Front Ecol Environ. 2012;10(3):135–143. https://doi.org/10.1890/110198
21. Meurisse N, Rassati D, Hurley BP, Brockerhoff EG, Haack RA. Common pathways by which non-native forest insects move internationally and domestically. J Pest Sci. 2019;92(1):13–27. https://doi.org/10.1007/s10340-018-0990-0
22. Hulcr J, Dunn RR. The sudden emergence of pathogenicity in insect-fungus symbioses threatens naive forest ecosystems. Proc R Soc B Biol Sci. 2011;278(1720):2866–2873. https://doi.org/10.1098/rspb.2011.1130
23. Allen E, Noseworthy M, Ormsby M. Phytosanitary measures to reduce the movement of forest pests with the international trade of wood products. Biol Invasions. 2017;19(11):3365–3376. https://doi.org/10.1007/s10530-017-1515-0
24. Renault D, Laparie M, McCauley SJ, Bonte D. Environmental adaptations, ecological filtering, and dispersal central to insect invasions. Annu Rev Entomol. 2018;63(1):345–368. https://doi.org/10.1146/annurev-ento-020117-043315
25. Cooperband MF, Stouthamer R, Carrillo D, Eskalen A, Thibault T, Cossé AA, et al. Biology of two members of the *Euwallacea fornicatus* species complex (Coleoptera: Curculionidae: Scolytinae), recently invasive in the U.S.A., reared on an ambrosia beetle artificial diet. Agric For Entomol. 2016;18(3):223–237. https://doi.org/10.1111/afe.12155

26. Huang J, Kautz M, Trowbridge AM, Hammerbacher A, Raffa KF, Adams HD, et al. Tree defence and bark beetles in a drying world: Carbon partitioning, functioning and modelling. New Phytol. 2020;225(1):26–36. https://doi.org/10.1111/nph.16173

27. Whattam M, Clover G, Firko M, Kalaris T. The biosecurity continuum and trade: Border operations. In: Gordh G, McKirdy S, editors. The handbook of plant biosecurity. Dordrecht: Springer; 2014. p. 149–188. https://doi.org/10.1007/978-94-007-7365-3_6

28. Eschen R, Rigaux L, Sukovata L, Vettraino AM, Marzano M, Grégoire JC. Phytosanitary inspection of woody plants for planting at European Union entry points: A practical enquiry. Biol Invasions. 2015;17(8):2403–2413. https://doi.org/10.1007/s10530-015-0883-6

29. Parliament of the Republic of South Africa. Plant health (phytosanitary) Bill. 2017 p. 1–24.

30. Haack RA, Petrice TR, Wiedenhoeft AC. Incidence of bark and woodboring insects in firewood: A survey at Michigan's Mackinac bridge. J Econ Entomol. 2010;103(5):1682–1692. https://doi.org/10.1603/ec10041

31. Liebhold AM, Brockerhoff EG, Kimberley M. Depletion of heterogeneous source species pools predicts future invasion rates. J Appl Ecol. 2017;54(6):1968–1977. https://doi.org/10.1111/1365-2664.12895

32. Grousset F, Grégoire JC, Jactel H, Battisti A, Beloglavec AB, Hrašovec B, et al. The risk of bark and ambrosia beetles associated with imported non-coniferous wood and potential horizontal phytosanitary measures. Forests. 2020;11(3):1–17. https://doi.org/10.3390/f11030342

33. Umeda C, Paine T. Temperature can limit the invasion range of the ambrosia beetle *Euwallacea* nr. *fornicatus*. Agric For Entomol. 2019;21(1):1–7. https://doi.org/10.1111/afe.12297

34. Peer K, Taborsky M. Outbreeding depression, but no inbreeding depression in haplodiploid ambrosia beetles with regular sibling mating. Evolution. 2005;59(2):317–323. https://doi.org/10.1111/j.0014-3820.2005.tb00992.x

35. Simberloff D. The role of propagule pressure in biological invasions. Annu Rev Ecol Evol Syst. 2009;40(1):81–102. https://doi.org/10.1146/annurev.ecolsys.110308.120304

36. Dodge C, Carrillo J, Eskalen A, Stouthamer R. Evidence for symbiont promiscuity in two invasive ambrosia beetles (Coleoptera: Scolytinae: *Euwallacea* spp.). Paper presented at: Annual Meeting of the Entomological Society of America; 2017 November 5–8; Denver, Colorado, USA. p. 165.

37. Freeman S, Miller G, Protasov A, Maymon M, Elazar M, David-Schwartz R, et al. Aposymbiotic interactions of three ambrosia beetle fungi with avocado trees. Fungal Ecol. 2019;39:117–130. https://doi.org/10.1016/J.FUNECO.2018.11.007

38. Lynch SC, Eskalen A, Gilbert GS. Host evolutionary relationships explain tree mortality caused by a generalist pest–pathogen complex. Evol Appl. 2021;14(4):1083-1094. https://doi.org/10.1111/eva.13182

39. Norris DM, Baker JM. A minimal nutritional substrate required by *Fusarium solani* to fulfill its mutualistic relationship with *Xyleborus ferrugineus*. Ann Entomol Soc Am. 1968;61(6):1473–1475. https://doi.org/10.1093/aesa/61.6.1473

40. Owens D, Seo M, Montgomery WS, Rivera MJ, Stelinski LL, Kendra PE. Dispersal behaviour of *Euwallacea* nr. *fornicatus* (Coleoptera: Curculionidae: Scolytinae) in avocado groves and estimation of lure sampling range. Agric For Entomol. 2019;21(2):199–208. https://doi.org/10.1111/afe.12321

41. Peel MC, Finlayson BL, McMahon TA. Updated world map of the Köppen-Geiger climate classification. Hydrol Earth Syst Sci Discuss. 2007;4(2):439–473. https://doi.org/10.5194/hess-11-1633-2007

42. Potgieter LJ, Douwes E, Gaertner M, Measey J, Paap T, Richardson DM. Biological invasions in South Africa's urban ecosystems: Patterns, processes, impacts, and management. In: Van Wilgen B, Measey J, Richardson DM, Wilson JR, Zengeya TA, editors. Biological Invasions in South Africa. New York: Springer International Publishing; 2020. p. 275–309. https://doi.org/10.1007/978-3-030-32394-3

43. Ploetz RC, Hulcr J, Wingfield MJ, De Beer ZWZ Destructive tree diseases associated with ambrosia and bark beetles: Black swan events in tree pathology? Plant Dis. 2013;97(7):856–872. https://doi.org/10.1094/PDIS-01-13-0056-FE

44. Paap T, Burgess TI, Wingfield MJ. Urban trees: Bridge-heads for forest pest invasions and sentinels for early detection. Biol Invasions. 2017;19(12):3515–3526. https://doi.org/10.1007/s10530-017-1595-x

45. Richardson DM, Pyšek P. Plant invasions: Merging the concepts of species invasiveness and community invasibility. Prog Phys Geogr. 2006;30:409–431. https://doi.org/10.1191/0309133306pp490pr

46. Sukopp H, Starfinger U. Disturbance in urban ecosystems. In: Walker LR, editor. Ecosystems of disturbed ground. Amsterdam: Elsevier; 1999. p. 397–412.

47. Kühnholz S, Borden JH, Uzunovic A. Secondary ambrosia beetles in apparently healthy trees: Adaptations, potential causes and suggested research. Integr Pest Manag Rev. 2003;6(3–4):209–219. https://doi.org/10.1023/A:1025702930580

48. Mendel Z, Protasov A, Maoz Y, Maymon M, Miller G, Elazar M, et al. The role of *Euwallacea* nr. *fornicatus* (Coleoptera: Scolytinae) in the wilt syndrome of avocado trees in Israel. Phytoparasitica. 2017;45(3):341–359. https://doi.org/10.1007/s12600-017-0598-6

49. Eskalen A, Gonzalez A, Wang DH, Twizeyimana M, Mayorquin JS, Lynch SC. First report of a *Fusarium* sp. and its vector tea shot hole borer (*Euwallacea fornicatus*) causing fusarium dieback on avocado in California. Plant Dis. 2012;96(7):1070. https://doi.org/10.1094/PDIS-03-12-0276-PDN

50. Coleman TW, Poloni AL, Chen Y, Thu PQ, Li Q, Sun J, et al. Hardwood injury and mortality associated with two shot hole borers, *Euwallacea* spp., in the invaded region of southern California, USA, and the native region of Southeast Asia. Ann For Sci. 2019;76(61):1–18. https://doi.org/10.1007/s13595-019-0847-6

51. Hulcr J, Black A, Prior K, Chen CY, Li HF. Studies of ambrosia beetles (Coleoptera: Curculionidae) in their native ranges help predict invasion impact. Florida Entomol. 2017;100(2):257–261. https://doi.org/10.1653/024.100.0219

52. De Wit MP, Crookes DJ, Blignaut JN, De Beer ZW, Paap T, Roets F, et al. Invasion of the polyphagous shot hole borer beetle in South Africa: A preliminary assessment of the economic impacts [document on the Internet]. c2021 [cited 2021 Jun 29]. Available from: https://doi.org/10.21203/rs.3.rs-220132/v1

53. Donovan GH, Butry DT, Michael YL, Prestemon JP, Liebhold AM, Gatziolis D, et al. The relationship between trees and human health. Am J Prev Med. 2013;44(2):139–145. https://doi.org/10.1016/j.amepre.2012.09.066

54. Boland JM. The impact of an invasive ambrosia beetle on the riparian habitats of the Tijuana River Valley, California. PeerJ. 2016;4(6), e2141. https://doi.org/10.7717/peerj.2141

55. Boland JM, Woodward DL. Impacts of the invasive shot hole borer (*Euwallacea kuroshio*) are linked to sewage pollution in southern California: The Enriched Tree Hypothesis. PeerJ. 2019;7, e6812. https://doi.org/10.7717/peerj.6812

56. Boland JM, Uyeda KA. The ecology and management of the Kuroshio Shot Hole Borer in the Tijuana River Valley [document on the Internet]. c2020 [cited 2021 Mar 25]. Available from: https://trnerr.org/wp-content/uploads/2020/05/KSHB-TRValley2020.pdf

57. Machingambi NM, Roux J, Dreyer LL, Roets F. Bark and ambrosia beetles (Curculionidae: Scolytinae), their phoretic mites (Acari) and associated *Geosmithia* species (Ascomycota: Hypocreales) from *Virgilia* trees in South Africa. Fungal Biol. 2014;118(5–6):472–483. https://doi.org/10.1016/j.funbio.2014.03.006

58. Coetsee C, Wigley BJ. *Virgilia divaricata* may facilitate forest expansion in the afrotemperate forests of the southern Cape, South Africa. Koedoe. 2013;55(1), Art. #1128. https://doi.org/10.4102/koedoe.v55i1.1128

59. Mayorquin JS, Carrillo JD, Twizeyimana M, Peacock BB, Sugino KY, Na F, et al. Chemical management of invasive shot hole borer and *Fusarium* dieback in California sycamore (*Platanus racemosa*) in southern California. Plant Dis. 2018;102(7):1307–1315. https://doi.org/10.1094/PDIS-10-17-1569-RE

60. Eatough Jones M, Paine TD. Potential pesticides for control of a recently introduced ambrosia beetle (*Euwallacea* sp.) in southern California. J Pest Sci. 2018;91(1):237–246. https://doi.org/10.1007/s10340-017-0866-8

61. Grosman DM, Eskalen A, Brownie C. Evaluation of emamectin benzoate and propiconazole for management of a new invasive shot hole borer (*Euwallacea* nr. *fornicatus*, Coleoptera: Curculionidae) and symbiotic fungi in California sycamores. J Econ Entomol. 2019;112(3):1267–1273. https://doi.org/10.1093/jee/toy423

62. California Forest Pest Council. California Forest Pest Conditions [document on the Internet]. c2021 [cited 2021 Feb 21]. Available from: https://www.fs.usda.gov/Internet/FSE_DOCUMENTS/fseprd896758.pdf

63. Jones ME, Kabashima J, Eskalen A, Dimson M, Mayorquin JS, Carrillo JD, et al. Evaluations of insecticides and fungicides for reducing attack rates of a new invasive ambrosia beetle (*Euwallacea* sp., Coleoptera: Curculionidae: Scolytinae) in infested landscape trees in California. J Econ Entomol. 2017;110(4):1611–1618. https://doi.org/10.1093/jee/tox163

64. Bonilla-Landa I, De la Cruz OL, Sanchéz-Rangel D, Ortíz-Castro R, Rodriguez-Haas B, Barrera-Méndez F, et al. Design, synthesis and biological evaluation of novel fungicides for the management of *Fusarium* dieback disease. J Mex Chem Soc. 2018;62(3):86–98. https://doi.org/10.29356/jmcs.v62i3.531

65. Van Der Werf HMG. Assessing the impact of pesticides on the environment. Agric Ecosyst Environ. 1996;60(2–3):81–96. https://doi.org/10.1016/S0167-8809(96)01096-1

66. Chen Y, Coleman TW, Poloni AL, Nelson L, Seybold SJ. Reproduction and control of the invasive polyphagous shot hole borer, *Euwallacea* nr. *fornicatus* (Coleoptera: Curculionidae: Scolytinae), in three species of hardwoods: Effective sanitation through felling and chipping. Environ Entomol. 2020;49(5):1155–1163. https://doi.org/10.1093/ee/nvaa103

67. Haack RA. Intercepted Scolytidae (Coleoptera) at US ports of entry: 1985–2000. Integr Pest Manag Rev. 2001;6:254–282. https://doi.org/10.1023/A:1025715200538

68. Van der Colff D, Dreyer LL, Valentine A, Roets F. Invasive plant species may serve as a biological corridor for the invertebrate fauna of naturally isolated hosts. J Insect Conserv. 2015;19(5):863–875. https://doi.org/10.1007/s10841-015-9804-3

69. Paap T, Wingfield MJ, De Beer ZW, Roets F. Lessons from a major pest invasion: The polyphagous shot hole borer in South Africa. S Afr J Sci. 2020;116(11/12), Art. #8757. https://doi.org/10.17159/sajs.2020/8757

**AUTHORS:**
Stacey Duvenage[1]
Werner Rossouw[1]
Germán Villamizar-Rodríguez[1]
Erika M. du Plessis[1]
Lise Korsten[1]

**AFFILIATION:**
[1]DSI–NRF Centre of Excellence in Food Security, Department of Plant and Soil Sciences, University of Pretoria

**CORRESPONDENCE TO:**
Lise Korsten

**EMAIL:**
lise.korsten@up.ac.za

**DATES:**
**Received:** 20 July 2020
**Revised:** 09 July 2021
**Accepted:** 23 July 2021
**Published:** 29 Nov. 2021

**HOW TO CITE:**
Duvenage S, Rossouw W, Villamizar-Rodríguez G, Du Plessis EM, Korsten L. Antibiotic resistance profiles of *Staphylococcus* spp. from white button mushrooms and handlers. S Afr J Sci. 2021;117(11/12), Art. #8667. https://doi.org/10.17159/sajs.2021/8667

**ARTICLE INCLUDES:**
☒ Peer review
☐ Supplementary material

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

**EDITORS:**
Pascal Bessong
Sandiswa Mbewana

**KEYWORDS:**
staphylococci, antibiotic genes, antibiotic resistance

**FUNDING:**
National Research Foundation (UID 74426), DSI–NRF Centre of Excellence in Food Security, University of Pretoria

# Antibiotic resistance profiles of *Staphylococcus* spp. from white button mushrooms and handlers

The presence of *Staphylococcus* spp. has increasingly been reported in food products and poses a public health threat. The aim of this study was to determine the diversity of *Staphylococcus* spp. and the antibiotic resistance profiles of isolates obtained from freshly harvested and packed ready-to-eat mushrooms ($n$=432) and handlers' hands ($n$=150). A total of 56 *Staphylococcus* isolates [46.4% ($n$=26) from hands and 53.6% ($n$=30) from mushrooms] were recovered belonging to 10 species. *Staphylococcus succinus* isolates ($n$=21) were the most prevalent, of which 52.4% came from mushrooms and 47.6% from hands. This was followed by *S. equorum* isolates [$n$=12; 91.7% ($n$=11) from mushrooms and 8.3% ($n$=1) from hands] and *S. saprophyticus* [$n$=9; 66.7% ($n$=6) from mushrooms and 33.3% ($n$=3) from hands]. Six isolates that were characterised as multidrug resistant were isolated from hands of handlers. Most (83.9%; $n$=47) of the 56 isolates were resistant to penicillin [53.2% ($n$=25) from mushrooms and 46.8% ($n$=22) from hands] and 14.3% ($n$=8) were resistant to cephalosporin classes [25% ($n$=2) from mushrooms and 75% ($n$=6) from hands], both of which are used to treat staphylococcal infections. Antibiotic resistance genes *blaZ* [25.0% ($n$=14) of all isolates of which 71.4% ($n$=10) were from hands and 28.57% ($n$=4) from mushrooms], *tetL* and *tetK* [both 1.8% ($n$=1) from hands], *mecA* [5.4% ($n$=3) from hands] and *ermA* [1.8% ($n$=1) from mushrooms] were detected from the 56 isolates. Only two (25.0%) of the eight methicillin-resistant staphylococci harboured the *mecA* gene, while only 11 (23%) of the 47 penicillin-resistant isolates harboured the *blaZ* gene [36.4% ($n$=4) from mushrooms and 63.6% ($n$=7) from hands]. Our results demonstrate that food handlers and harvested and packed ready-to-eat mushrooms could be a source of diverse *Staphylococcus* spp. that exhibit antimicrobial resistance. Clinically relevant *S. aureus* was only detected on one handler's hand; however, the isolate was not multidrug resistant. The presence of diverse *Staphylococcus* spp. on mushrooms and the hands of handlers is a potential public health concern due to their potential to cause opportunistic infections.

**Significance:**

- This study is the first to describe the antibiotic resistance profiles and antibiotic gene presence of *Staphylococcus* spp. isolated from fresh mushrooms and hands of pickers and packers. Mushrooms and handlers in this study were demonstrated to be possible routes of transmission of *Staphylococcus* spp. that are antibiotic resistant and which harbour antibiotic resistance genes, presenting a possible public health hazard.

## Introduction

*Staphylococcus* spp. are ubiquitous and transient organisms. *Staphylococcus aureus* is known to be a natural coloniser of human skin, and between 10% and 20% of adults' skin is persistently colonised by *S. aureus,* while between 30% and 50% of healthy people's skin is colonised by *S. aureus*.[1,2] *Staphylococcus aureus* specifically has been recorded to cause invasive infections or toxin-mediated diseases, including endocarditis, bacteraemia, metastatic infections and toxic shock syndrome.[3] Traditionally, much attention has been paid to *S. aureus* as an organism causing infection; however, coagulase-negative *Staphylococcus* spp. have more recently been shown to also be pathogenic.[4]

Treatment of staphylococcal infections with antibiotics has become common practice in the medical field and, subsequently, higher levels of resistance have been recorded.[2,5] The spread of antimicrobial-resistant staphylococci represents a hazard to human and veterinary health[6] because staphylococci have the ability to transfer antibiotic resistance genes (ARGs) to other pathogenic organisms[2,5]. Previous studies have also shown that staphylococci can be important reservoirs of ARGs in ready-to-eat food.[7,8] Moreover, the exchange of genetic material, such as mobile elements, has reportedly been associated with food-processing environments.[9]

Multidrug-resistant (MDR) staphylococci add to the public health concerns with respect to staphylococcal infections. For example, it has been associated with an increased severity of infections as well as a growing number of people being hospitalised due to such infections.[10] In the United States of America, 60% of infectious disease specialists reported that untreatable bacterial infections had been observed, highlighting the impact on human health.[11] The prevalence of MDR organisms in environmental samples has further heightened the concern of staphylococcal infections for the World Health Organization[10] due to the potential horizontal transfer of ARGs, genetic mutation and recombination. Food associated with bacteria harbouring such ARGs is a major concern and represents possible reservoirs for the spread of these ARGs.[6]

Food commodities that are handled extensively and do not go through a decontamination step might therefore harbour antimicrobial-resistant microorganisms and thus pose a human and environmental health threat. Food products

harbouring antimicrobial-resistant *Escherichia coli*, *Campylobacter* spp., *Salmonella* spp., *Clostridium* spp. and *Listeria monocytogenes* have previously been documented and there has now been an increased interest in the role of *Staphylococcus* spp. isolated from food.[2,6] Therefore, the aim of this study was to determine the diversity of *Staphylococcus* spp. as well as the antibiotic resistance (AR) profiles and the presence of ARGs in isolates obtained from mushrooms and mushroom handlers. Mushrooms that are handled extensively by pickers and packers represent a potential risk if hygiene principles are not observed. This study is the first of its kind to investigate the diversity and AR profiles of *Staphylococcus* spp. on mushrooms and handlers' hands.

## Materials and methods

A total of 432 white button mushroom samples were collected as outlined in Rossouw and Korsten[12]. Sampling sites included two large-scale commercial mushroom farms located in Gauteng Province, South Africa, which follow similar production practices, operating under Global-G.A.P. Integrated Farm Assurance Standards V5.1. Mushrooms that were void of defects and at the ready-to-harvest stage were randomly sampled based on uniformity of size, shape and maturity. Mushrooms were aseptically harvested by researchers.[12] Packed mushrooms handled by pickers and packers prior to punnet sealing were also collected. Sampling of hands was done according to standard hand swab procedures.[13] Hand swabs were collected using the Copan Venturi Transystem (Copan, Italy) from five pickers' and five packers' dominant hands on a weekly basis for a period of 15 weeks, resulting in 150 hand samples. Samples were placed in a cooler box and transported to the laboratory for analysis within 24 h. At the time of the study, ethical clearance for non-invasive swabbing of the hands was not mandatory and was therefore waived by the Ethics Committee of the Faculty of Natural and Agricultural Sciences, University of Pretoria. Commercial and personal permission to conduct non-invasive hand swabbing was granted by the commercial entity as well as by the persons whose hands were sampled. Confidentiality was maintained throughout the process.

A sample (250 g) was aseptically obtained from each mushroom sample and homogenised using a handheld blender (Russell Hobbs, Johannesburg, South Africa). Ten grams of the homogenised sample was added to 90 mL tryptone soy broth (Merck, Johannesburg) in a sterile homogeniser bag, macerated for 5 min and incubated for 24 h at 37 °C. Contents were subsequently plated onto Baird-Parker agar (Merck) and incubated for 24 h at 37 °C.

Hand swabs were placed into 9 mL buffered peptone water (Merck), and incubated for 24 h at 37 °C and plated onto Baird-Parker agar plates, which were incubated for 24 h at 37 °C. Presumptive *Staphylococcus* spp. were selected from the Baird-Parker agar, based on a dark grey to black colony morphology which was surrounded by a clear zone.[14] Isolates were purified on Baird-Parker agar and the purified isolate identity was determined using matrix-assisted laser desorption ionisation time-of-flight mass spectrometry in combination with the Bruker Biotyper software and database.[15]

A total of 56 purified and confirmed *Staphylococcus* spp. isolates were used for further characterisation of AR, using the Kirby–Bauer disk diffusion method.[16] Each isolate was cultured in 9 mL of brain heart infusion broth (Merck) and incubated for 18 h at 37 °C (which resulted in approximately log 8 cfu/mL[17]) and subsequently plated onto Mueller–Hinton agar plates (Merck). The Kirby–Bauer disk diffusion test was employed to determine susceptibility to the antibiotics listed in Table 1. Zone diameters were measured and interpreted according to the Clinical and Laboratory Standards Institute guidelines.[17] Strains resistant to three or more antibiotic classes were defined as MDR.

**Table 1:** Antibiotics tested using the Kirby–Bauer disk diffusion method

| Antibiotic class | Antibiotic | Concentration |
|---|---|---|
| Cephalosporin | Cefoxitin | 30 μg |
| Phenicol | Chloramphenicol | 30 μg |
| Quinolone | Ciprofloxacin | 5 μg |
| Lincomycin | Clindamycin | 2 μg |
| Macrolide | Erythromycin | 15 μg |
| Aminoglycoside | Gentamicin | 10 μg |
| Penicillin | Oxacillin | 1 μg |
| | Penicillin | 10 μg |
| Rifamycin | Rifampicin | 5 μg |
| Tetracycline | Oxytetracycline | 30 μg |
| Glycopeptide | Teicoplanin | 30 μg |
| | Vancomycin | 30 μg |

**Table 2:** Summary of antibiotic resistance genes, associated primers and cycling conditions used for screening purposes

| Resistance | Gene | Primers | | Amplicon (annealing temperature) | Primer concentration | Reference | Control |
|---|---|---|---|---|---|---|---|
| Penicillin | *blaZ* | **Forward**<br>**Reverse** | **blaZF** (5'-ACT TCA ACA CCT GCT GCT TTC-3')<br>**blaZR** (5'-TGA CCA CTT TTA TCA GCA ACC-3') | 173 bp (55 °C) | 1.0 μM | 28 | *Staphylococcus aureus* ATCC43300 |
| Methicillin/ Oxacillin | mecA | **Forward**<br>**Reverse** | **mecA1** (5'-GGG ATC ATA GCG TCA TTA TTC-3')<br>**mecA2** (5'-AAC GAT TGT GAC ACG ATA GCC-3') | 527 bp (50 °C) | 0.2 μM | 20 | *Staphylococcus aureus* ATCC43300 |
| Tetracycline | *tetM* | **Forward**<br>**Reverse** | **tetMF** (5'-AGT GGA GCG ATT ACA GAA-3')<br>**tetMR** (5'-CAT ATG TCC TGG CGT GTC TA-3') | 158 bp (48 °C) | 0.2 μM | 29 | None |
| Tetracycline | *tetK* | **Forward**<br>**Reverse** | **tetKF** (5'-GTA GCG ACA ATA GGT AAT AGT-3')<br>**tetKR** (5'-GTA GTG ACA ATA AAC CTC CTA-3') | 460 bp (55 °C) | 0.2 μM | 29 | None |
| Tetracycline | *tetL* | **Forward**<br>**Reverse** | **tetLF** (5'-TGG TGG AAT GAT AGC CCA TT-3')<br>**tetLR** (5'-CAG GAA TGA CAG CAC GCT AA-3') | 229 bp (51 °C) | 0.2 μM | 30 | None |
| Erythromycin | *ermA* | **Forward**<br>**Reverse** | **ermA1** (5'-AAG CGG TAA AAC CCC TCT GAG-3')<br>**ermA2** (5'-TCA AAG CCT GTC GGA ATT GG-3') | 442 bp (52 °C) | 0.4 μM | 31 | None |
| Erythromycin | *ermC* | **Forward**<br>**Reverse** | **ermC1** (5'-ATC TTT GAA ATC GGC TCA GG-3')<br>**ermC2** (5'-CAA ACC CGT ATT CCA CGA TT-3') | 295 bp (50 °C) | 0.4 μM | 31 | None |
| Erythromycin | *ermB* | **Forward**<br>**Reverse** | **ermBF** (5'-TGG TAT TCC AAA TGC GTA ATG-3')<br>**ermBR** (5'-CTG TGG TAT GGC GGG TAA GT-3') | 745 bp (55 °C) | 0.4 μM | 30 | None |
| Cefoxitin | *ampC* | **Forward**<br>**Reverse** | **ampCF** (5'-GTG ACC AGA TAT GGC CAC A-3')<br>**ampCR** (5'-TTA CTG TAG CGC CTC GAG GA-3') | 822 bp (59 °C) | 0.6 μM | 32 | *Enterobacter cloacae* NCTC 13406 |
| 16S rRNA amplification control | | **Forward**<br>**Reverse** | **27F** (5'-GAG TTT GAT CCT GGC TCA G-3')<br>**1492R** (5'-TAC GGY TAC CTT GTT ACG ACT T-3') | 1465 bp | 0.4 μM | 33 | N/A |

Genomic DNA (gDNA) was extracted using the Quick-gDNA Miniprep kit (Zymo Research, Irvine, CA, USA), following overnight incubation in tryptone soy broth at 37 °C. The DNA concentration was measured using the broad range kit for Qubit 2.0 fluorometer (Life Technologies, Johannesburg, South Africa). Polymerase chain reaction (PCR) mixtures with a final volume of 25 $\mu$L, containing 10–100 $\mu$g gDNA, were prepared using Dream Taq PCR Master Mix (1x) (Thermo Scientific, Johannesburg, South Africa). Each reaction included 16S universal primers as amplification controls (Table 2) as well as the specific ARG primer with primer concentration as outlined in Table 2. PCR cycling conditions were as follows: 94 °C for 2 min, followed by 35 cycles of denaturation (30 s at 95 °C), annealing (30 s at primer temperature, Table 1) and extension (60 s at 72 °C), with a final extension at 72 °C for 5 min. PCR reactions were performed on a BioRad T100 thermocycler (BioRad, Johannesburg, South Africa). Amplicons were visualised on a 2% agarose gel stained with Roti®-safe (Carl Roth GmbH & Co, Karlsruhe, Germany) using a molecular imager in conjunction with the Image Lab™ software (BioRad).

## Results

Out of 582 samples collected, 56 (9.6%) were positive for *Staphylococcus* spp. All *Staphylococcus* spp. isolates and their AR and ARG profiles are presented in Figure 1 by sample and year collected. Ten *Staphylococcus* species were isolated: *S. aureus, S. epidermidis*, *S. equorum*, *S. haemolyticus, S. hominis, S. saprophyticus, S. sciuri, S. succinus, S. warneri* and *S. xylosus* (Figures 1 and 2). *Staphylococcus aureus* was only isolated from one packer's hand sample and no mushroom sample yielded *S. aureus*. *Staphylococcus epidermidis* was detected from one picker's and one packer's hand, and both isolates were found to be MDR. A total of 12 *S. equorum* isolates were detected and were isolated mainly from hands (91.67%; $n$=11/12) of pickers (41.67%; $n$=5/12) and packers (50%; $n$=6/12), with only 8.33% ($n$=1/12) isolated from a packed mushroom sample. Two MDR *S. equorum* isolates from pickers' hands, isolated from the same farm and same mushroom picking room (data not shown), shared a phenotypic MDR antibiotic profile but the ARGs were different. Only one MDR *S. haemolyticus* was isolated, from a packer's hand. One packer's hand and one packed mushroom sample were contaminated with *S. hominis*, with the one on the packed mushrooms found to be MDR. Mushrooms sampled before harvest ($n$=6) and three pickers' hands were contaminated with *S. saprophyticus*. Only one *S. sciuri* was isolated from one packed mushroom sample. *Staphylococcus succinus* was isolated from hands ($n$=10) as well as from mushrooms before picking ($n$=4) and after packing ($n$=7), with one found to be methicillin resistant due to the presence of *mecA*. Four of the six *S. succinus* isolates were isolated from pickers in the same growing room during the same sampling period (data not shown). *Staphylococcus xylosus* was only isolated from mushrooms sampled before picking ($n$=4) and after packing ($n$=2) – these isolates were resistant to only the penicillin class of antibiotics. *Staphylococcus warneri* was isolated from a packer's hand.

Six *Staphylococcus* spp. (10.7%; $n$=6/56) were considered MDR, with all isolated from handlers' hands. Expressed resistance to at least one antibiotic agent was found in 85.7% [$n$=48; 52.1% ($n$=25) from mushrooms and 47.9% ($n$=23) from hands] of all *Staphylococcus* spp. characterised. A total of 83.9% of the 56 isolates [$n$=47; 53.2% ($n$=25) from mushrooms and 46.8% ($n$=22) from hands] were resistant to the penicillin class. Resistance to both penicillin and oxacillin was observed in 32 of the 56 isolates, of which 46.9% ($n$=15) were from mushrooms and 53.1% ($n$=17) were from hands.

Of the 47 isolates resistant to the penicillin class, 23.4% ($n$=11) harboured the *blaZ* gene and 6.4% ($n$=3) harboured the *blaZ* gene but did not express resistance to penicillin nor oxacillin. Of the 11 that harboured the *blaZ* gene, 36.4% ($n$=4) were from mushrooms and 63.6% ($n$=7) were from hands. Of the 32 isolates resistant to oxacillin, 9.4% ($n$=3) of isolates from hands harboured the *mecA* gene. Three isolates from hands were resistant to oxytetracycline, with one isolate from hands harbouring *tetK* and one isolate from hands harbouring *tetL*. Only one mushroom isolate harboured *ermA.* No *ampC* was detected, even though eight *Staphylococcus* spp. were resistant to cefoxitin. Two of the three isolates from hands harbouring *mecA* were resistant to cefoxitin.

## Discussion

The status of antibiotic-resistant microorganisms in the agricultural environment is becoming a major concern in the global food industry and in the public health sector.[5] The prevalence and status of staphylococci in animal products like meat and cheese have been well established.[5,8,14,18-20]However, there is limited information on the prevalence of AR and ARG in *Staphylococcus* spp. isolated from fresh produce. This study, the first of its kind as far as we could determine, investigated the diversity and resistance of *Staphylococcus* spp. isolated from white button mushrooms and mushroom handlers.

Of the ten *Staphylococcus* species identified, seven were found to be associated with raw mushrooms – these were *S. epidermidis, S. equorum, S. hominis, S. saprophyticus, S. sciuri, S. succinus* and *S. xylosus*. Both *S. aureus* and *S. epidermidis* are associated with the human skin microflora; these as well as other staphylococci pose an important health concern, in addition to exhibiting AR. The presence of these species associated mainly with the hands of workers indicates the need for improved personal hygiene implementation, training and enforcement.[2] These isolates can contribute to illness of both the handlers as well as susceptible consumers.

In this study, *S. saprophyticus*, *S. succinus* and *S. xylosus* were mainly detected from mushrooms. A previous study determined the presence of *S. xylosus, S. epidermidis* and *S. saprophyticus* in ready-to-eat products from animal origin. Furthermore, these species are also commonly associated with farm animals.[8] Mushrooms are cultivated on compost composed of hay, chicken manure, leachate and agricultural lime, which are stored on farms for months in bulk, prior to composting. Antibiotic-resistant bacteria have been found to be present in the excreta of broiler chickens.[21] Graham et al.[21] concluded that typical storage of chicken manure was not sufficient to eliminate the antibiotic-resistant *Staphylococcus* spp. Compost used for the production of mushrooms, which includes chicken manure, is pasteurised at between 60 °C and 75 °C for 13 days. However, Fontes et al.[19] found that staphylococci were able to survive high temperature pasteurisation. Therefore, the presence of these species is not necessarily eliminated from mushroom compost during the pasteurisation process. In addition, previous research has demonstrated that the use of antibiotics in the chicken rearing industry has been linked to the presence of antibiotic-resistant organisms present on farm workers as well as the growing environment.[23] Moreover, flies have previously been reported to increase the human exposure to antibiotic-resistant bacteria.[23] The presence of *Staphylococcus* spp. could be due to survival of organisms during storage and pasteurisation because of the organism's innate ability to survive such conditions as well as the possibility of post-pasteurisation contamination. All these factors might lead to the establishment and spread of *Staphylococcus* species, and antibiotic-resistant and MDR staphylococci in the environment and food system.

There has been an increase in the number of antibiotic-resistant organisms associated with humans[2], their direct environments[20,22,24] and their food[2,8,14,18,19,25]. Transmission of AR in bacteria is further aided by the ability of food to be a vehicle.[8] In the current study, only 10% of all staphylococci isolates were considered MDR. Benjelloun Touimi et al.[2] found that 100% of *Staphylococcus* isolates from vegetable and food handlers were MDR. Moreover, in ready-to-eat food, 94.12%[25], 89.81%[22] and 32.8%[8] of *Staphylococcus* isolates investigated were found to be MDR, reaffirming that food can be a vehicle for the spread of MDR *Staphylococcus* spp.

Penicillin, oxacillin and methicillin are the first line of defence against clinical staphylococci infections. In this study, 83.9% of isolates were resistant to the penicillin class, with 14 isolates harbouring *blaZ*, the gene that encodes $\beta$-lactamase and confers resistance to penicillin. Similarly, 94% of *Staphylococcus* spp. isolated from chicken and beef[14], 78.5% from soft cheeses[19] and 67.4% from cow mastitis[20] were resistant to penicillin. Moreover, Benjelloun Touimi et al.[2] found that all *Staphylococcus* spp. isolates from ready-to-eat foods were resistant to penicillin and oxacillin. Klimiene et al.[20], however, found a 66% presence of the *blaZ* gene in staphylococci assessed, compared with our study's 25% *blaZ* gene prevalence.

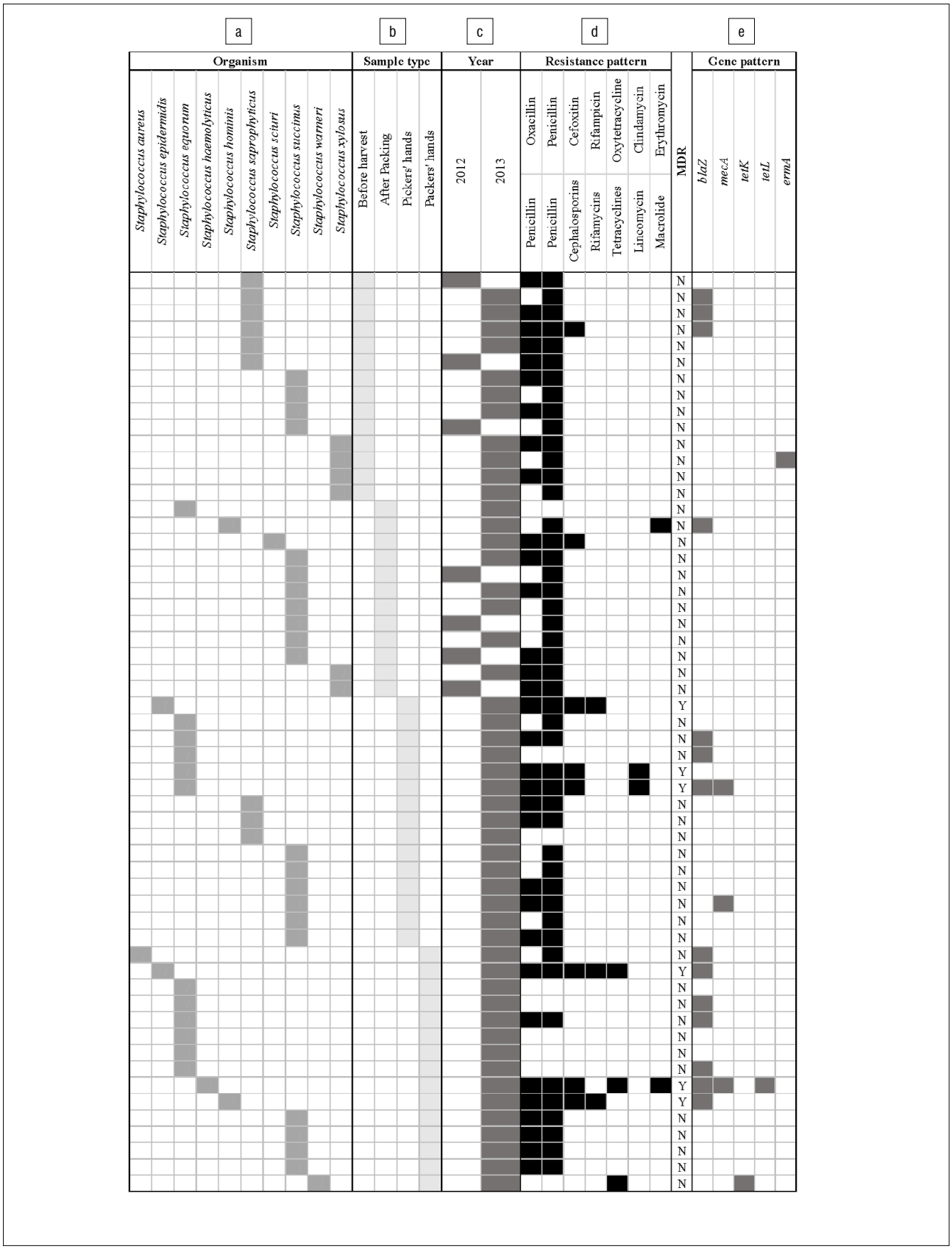



**Figure 1:** *Staphylococcus* species, phenotypic antibiotic susceptibility profile and antibiotic resistance gene profiles. Each line represents one isolate. (a) Shading indicates to which species the isolate belongs. (b) Shading indicates the isolation source. (c) Shading indicates the year of collection. (d) Shading indicates the resistance to an antibiotic. Only antibiotics for which resistance was determined are included and only antibiotic resistance genes present are listed. (e) Shading indicates the presence of a specified gene.

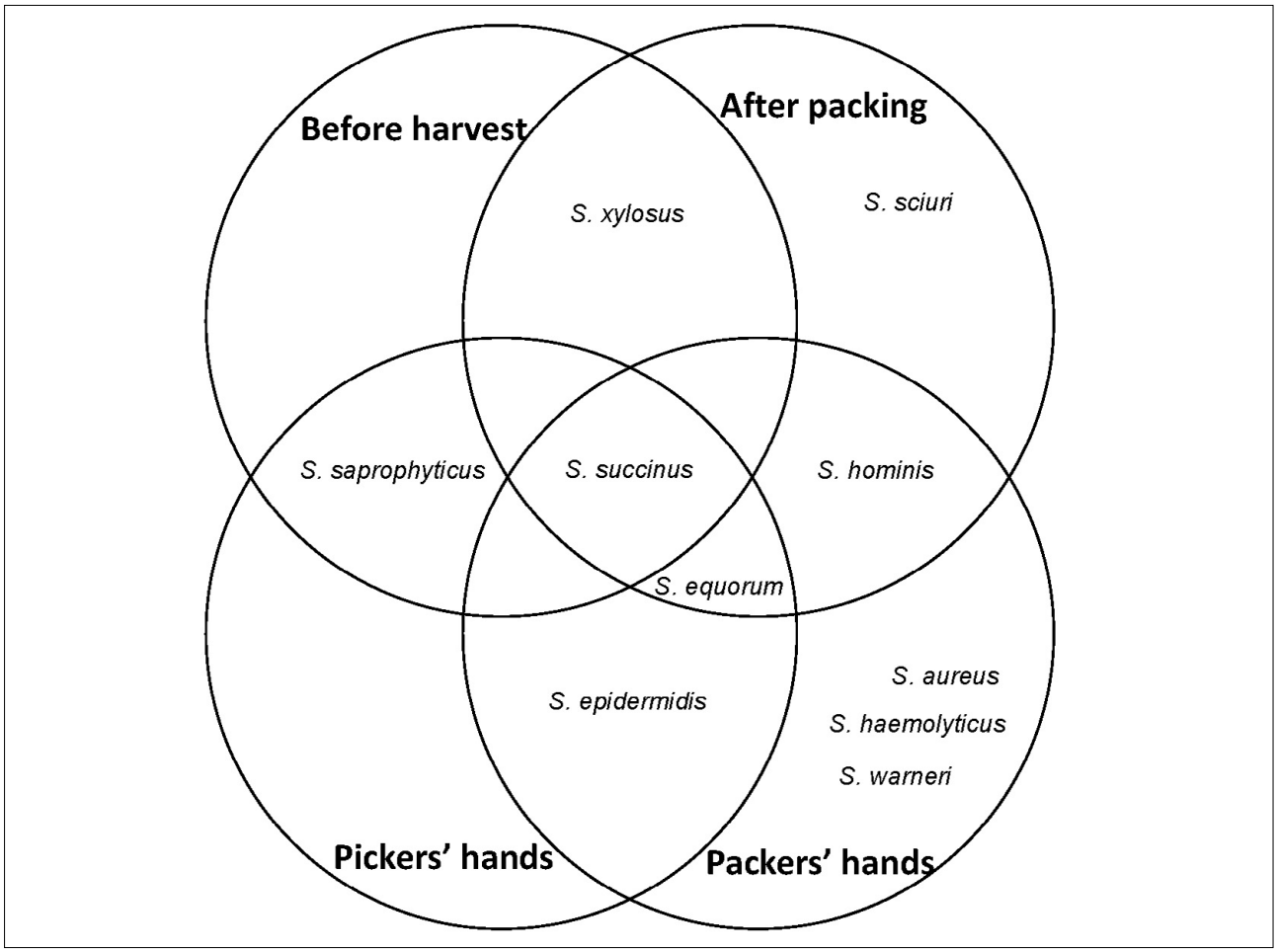



**Figure 2:** *Staphylococcus* species diversity indicating unique and shared species isolated from ready-to-eat mushrooms at harvest and mushrooms after packing as well as from the hands of pickers and packers.

The resistance of strains to cefoxitin is an indication of methicillin resistance in *Staphylococcus* phenotypes which are resistant to all $\beta$-lactam antibiotics, including penicillin and isoxazolyl penicillins.[8] The *mecA* gene confers resistance to methicillin (oxacillin and /or cefoxitin) and its potential transfer between organisms represents a public health concern.[2] In this study, 14.3% ($n$=8/56) of isolates were resistant to cefoxitin, of which 25.0% ($n$=2/8) were found on mushrooms and 75.0% ($n$=6/8) were found on hands; however, only two of the cefoxitin-resistant isolates harboured the *mecA* gene. Vyletelova et al.[26VLKOVA H., MANGA I. (2011] found that not all isolates (174/200) that showed resistance to cefoxitin harboured *mecA.* Chajęcka-Wierzchowska et al.[8] found that all but one isolate exhibiting cefoxitin resistance harboured *mecA*, which is part of the staphylococcal cassette chromosome *mec* (SCCmec) on the bacterial chromosome. All isolates in the current study that were resistant to cefoxitin were also resistant to oxacillin and penicillin. The co-occurrence of resistance to methicillin, oxacillin and penicillin has been described previously.[26] In addition, Chajęcka-Wierzchowska et al.[8] reported that isolates with resistance to methicillin, oxacillin and penicillin were also resistant to rifampicin, clindamycin and tetracycline. In this study, 33.3% of the eight methicillin-resistant staphylococci were also resistant to rifampicin, 25% to tetracycline and clindamycin and 25% to erythromycin.

Two mechanisms of tetracycline resistance have been identified in *Staphylococcus* spp. and are mediated by *tet* genes: plasmid-mediated *tetK* and *tetL* encoding efflux and *tetM* encoding ribosomal protection mediated determinants.[27] In the current study, two isolates harbouring *tetL* and *tetK* were found to be resistant to oxytetracycline; in addition, one isolate was found to be resistant to oxytetracycline without the presence of any of the *tet* genes tested for. Previous studies have found discrepancies between phenotypic expression and the presence of resistance genes.[27] In comparison, AR to tetracycline was seen in 14.7% of staphylococci isolates from soft cheese[19], 18.9% of staphylococci isolates from cow mastitis[20] and 34.5% of isolates from food of animal origin, of which all isolates resistant harboured at least one *tet* gene[8]. Osman et al.[14] found 68% AR to tetracycline from raw beef and chicken meat in Egypt, which was considerably higher than this and other studies. A conjugative transposon (Tn6079) is responsible for the spread of *tetL*[27], whereas *tetM* can be transposon-located or chromosomal[27]. Therefore, the presence of *tetL* and *tetM* can indicate the potential for horizontal gene transfer between organisms.

## Conclusion

In conclusion, a diverse number of *Staphylococcus* spp. were associated with mushrooms and mushroom handlers' hands. Antibiotic resistance of these mushroom- and hand-associated *Staphylococcus* spp. demonstrates a public health threat due to the potential of antibiotic gene transfer to medically important *Staphylococcus* spp. Moreover, opportunistic infection that might result due to an AR *Staphylococcus* spp. could lead to an infection that is difficult to treat. The presence of AR organisms adds to the general concern around the reservoir of AR on food products and food handlers' hands. Future research should determine the source of *Staphylococcus* spp. in production and on the product at the market end, in order to determine the specific risk to the final consumer. In addition, future research should also focus on ARG transfer and mechanisms in *Staphylococcus* species within the agricultural environment.

## Acknowledgements

We thank Ms L. Pattyn for PCR analysis. The MALDI-TOF equipment was based on the research supported in part by the National Research Foundation (NRF) of South Africa (grant specific unique reference number (UID) 74426). The financial assistance of the NRF towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the authors and are not necessarily to be attributed to the NRF. S.D. and W.R. acknowledge the Department of Science and Innovation–NRF Centre of Excellence in Food Security for financial assistance. G.V.R. acknowledges the University of Pretoria for financial assistance.

## Competing interests

We have no competing interests to declare.

## Authors' contributions

S.D.: Conceptualisation, research design and interpretation of data (conceptualisation), laboratory work, data collection and analysis (methodology and data collection), framework, final revision of manuscript, editing for intellectual content and submission (drafting or critically revising the manuscript), project management. W.R.: Research design (conceptualisation), field work, sampling, laboratory work, data collection and analysis (methodology and data collection), framework and writing of manuscript (drafting or critically revising the manuscript), project management. G.V.R.: Conceptualisation, research design and interpretation of data (conceptualisation), laboratory work, data collection and analysis (methodology and data collection), framework, final revision of manuscript, editing for intellectual content (critically revising the manuscript). E.d.P.: Conceptualisation, research design and interpretation of data (conceptualisation), data analysis (methodology and data collection), framework, final revision of manuscript, editing for intellectual content and submission (drafting or critically revising the manuscript). L.K.: Research design (conceptualisation), data analysis (methodology and data collection), framework, final revision of manuscript, editing for intellectual content and submission (drafting or critically revising the manuscript), student supervision, project leadership, funding acquisition.

## References

1. Noble W, Valkenburg H, Wolters C. Carriage of *Staphylococcus aureus* in random samples of a normal population. Epidemiol Infect. 1967;65(4):567–573. https://doi.org/10.1017/S002217240004609X
2. Benjelloun Touimi G, Bennani L, Berrada S, Moussa B, Bennani B. Prevalence and antibiotic resistance profiles of *Staphylococcus* sp. isolated from food, food contact surfaces and food handlers in a Moroccan hospital kitchen. Lett Appl Microbiol. 2020;70(4):241–251. https://doi.org/10.1111/lam.13278
3. Lowy F. *Staphylococcus aureus* infections. N Engl J Med. 1998;339(8):520–532. https://doi.org/10.1056/NEJM199808203390806
4. Gillespie B, Headrick S, Boonyayatra S, Oliver S. Prevalence and persistence of coagulase-negative *Staphylococcus* species in three dairy research herds. Vet Microbiol. 2009;134(1–2):65–72. https://doi.org/10.1016/j.vetmic.2008.09.007
5. Osman K, Alvarez-Ordóñez A, Ruiz L, Badr J, ElHofy F, Al-Maary KS, et al. Antimicrobial resistance and virulence characterization of *Staphylococcus aureus* and coagulase-negative staphylococci from imported beef meat. Ann Clin Microbiol Antimicrob. 2017;16(1):35. https://doi.org/10.1186/s12941-017-0210-4
6. Chajecka-Wierzchowska W, Zadernowska A, Nalepa B, Sierpi´Nska M, Łaniewska-Trokenheim Ł. Retail ready-to-eat food as a potential vehicle for *Staphylococcus* spp. harboring antibiotic resistance genes. J Food Prot. 2014;77(6):993–998. https://doi.org/10.4315/0362-028X.JFP-13-466
7. Podkowik M, Bystroń J, Bania J. Prevalence of antibiotic resistance genes in staphylococci isolated from ready-to-eat meat products. Pol J Vet Sci. 2012;15(2):233–237. https://doi.org/10.2478/v10181-011-0139-z
8. Chajęcka-Wierzchowska W, Zadernowska A, Nalepa B, Sierpinska M, Laniewska-Trokenheim L. Coagulase-negative staphylococci (CoNS) isolated from ready-to-eat food of animal origin – phenotypic and genotypic antibiotic resistance. Food Microbiol. 2015;46:222–226. https://doi.org/10.1016/j.fm.2014.08.001
9. Virdis S, Scarano C, Cossu F, Spanu V, Spanu C, De Santis EPL. Antibiotic resistance in *Staphylococcus aureus* and coagulase negative staphylococci isolated from goats with subclinical mastitis. Vet Med Int. 2010;2010:1–6. https://doi.org/10.4061/2010/517060
10. World Health Organization. Antimicrobial resistance: Global report on surveillance. Geneva: World Health Organization; 2014. Available from: http://apps.who.int/iris/bitstream/10665/112642/1/9789241564748_eng.pdf
11. Spellberg B, Gilbert DN. The future of antibiotics and resistance: A tribute to a career of leadership by John Bartlett. Clin Infect Dis. 2014;59:S71–S75. https://doi.org/10.1093/cid/ciu392
12. Rossouw W, Korsten L. Cultivable microbiome of fresh white button mushrooms. Lett Appl Microbiol. 2017;64(2):164–170. https://doi.org/10.1111/lam.12698
13. ISO 18593. Microbiology of food and animal feeding stuffs – Horizontal methods for sampling techniques from surfaces using contact plates and swabs. Iso 18593 [webpage on the Internet]. c2004 [cited 2021 Jul 19]. Available from: http://www.iso.org/iso/catalogue_detail.htm?csnumber=39849
14. Osman KM, Amer AM, Badr JM, Saad ASA. Prevalence and antimicrobial resistance profile of *Staphylococcus* species in chicken and beef raw meat in egypt. Foodborne Pathog Dis. 2015;12(5):406–413. https://doi.org/10.1089/fpd.2014.1882
15. Standing T-A, Du Plessis E, Duvenage S, Korsten L. Internalisation potential of *Escherichia coli* O157:H7, *Listeria monocytogenes*, *Salmonella enterica* subsp. *enterica* serovar Typhimurium and *Staphylococcus aureus* in lettuce seedlings and mature plants. J Water Health. 2013;11(2):210. https://doi.org/10.2166/wh.2013.164
16. Bauer A, Kirby W, Sherris J, Turck M. Antibiotic susceptibility testing by a standardized single disk method. Am J Clin Pathol. 1966;45(4):493–496. https://doi.org/10.1093/ajcp/45.4_ts.493
17. Clinical and Laboratory Standards Institute (CLSI). Performance standards for antimicrobial susceptibility testing. CLSI supplement M100. Annapolis Junction, MD: CLSI; 2018.
18. Nunes RSC, Pires de Souza C, Pereira KS, Del Aguila EM, Flosi Paschoalin VM. Identification and molecular phylogeny of coagulase-negative staphylococci isolates from Minas Frescal cheese in southeastern Brazil: Superantigenic toxin production and antibiotic resistance. J Dairy Sci. 2016;99(4):2641–2653. https://doi.org/10.3168/jds.2015-9693
19. Fontes CO, Silva VL, De Paiva MRB, Garcia RA, Resende JA, Ferreira-Machado AB, et al. Prevalence, antimicrobial resistance, and virulence characteristics of mecA-encoding coagulase-negative staphylococci isolated from soft cheese in Brazil. J Food Sci. 2013;78(4):594–599. https://doi.org/10.1111/1750-3841.12088
20. Klimiene I, Virgailis M, Pavilonis A, Siugzdiniene R, Mockeliunas R, Ruzauskas M. Phenotypical and genotypical antimicrobial resistance of coagulase-negative staphylococci isolated from cow mastitis. Pol J Vet Sci. 2016;19(3):639–646. https://doi.org/10.1515/pjvs-2016-0080
21. Graham JP, Evans SL, Price LB, Silbergeld EK. Fate of antimicrobial-resistant enterococci and staphylococci and resistance determinants in stored poultry litter. Environ Res. 2009;109(6):682–689. https://doi.org/10.1016/j.envres.2009.05.005
22. Wu S, Huang J, Zhang F, Wu Q, Zhang J, Pang R, et al. Prevalence and characterization of food-related methicillin-resistant *Staphylococcus aureus* (MRSA) in China. Front Microbiol. 2019;10:1–13. https://doi.org/10.3389/fmicb.2019.00304
23. Graham JP, Price LB, Evans SL, Graczyk TK, Silbergeld EK. Antibiotic resistant enterococci and staphylococci isolated from flies collected near confined poultry feeding operations. Sci Total Environ. 2009;407(8):2701–2710. https://doi.org/10.1016/j.scitotenv.2008.11.056

24. Papadopoulos P, Papadopoulos T, Angelidis AS, Kotzamanidis C, Zdragas A, Papa A, et al. Prevalence, antimicrobial susceptibility and characterization of *Staphylococcus aureus* and methicillin-resistant *Staphylococcus aureus* isolated from dairy industries in north-central and north-eastern Greece. Int J Food Microbiol. 2019;291:35–41. https://doi.org/10.1016/j.ijfoodmicro.2018.11.007

25. Bunnueang N, Kongpheng S, Singkhamanan K, Saengsuwan P, Rattanachuay P, Dangsriwan S, et al. Methicillin-resistant *Staphylococcus aureus* from ready-to-eat foods in a hospital canteen, southern Thailand: Virulence characterization and genetic relationship. Southeast Asian J Trop Med Public Health. 2015;46(1):86–96.

26. Vyletelova M, Vlkova H, Manga I. Occurrence and characteristics of methicillin resistant *Staphylococcus aureus* and methicillin resistant coagulase-negative staphylococci in raw milk manufacturing. Czech J Food Sci. 2011;29:S11–S16. https://doi.org/10.17221/4443-CJFS

27. Trzcinski K, Cooper BS, Hryniewicz W, Dowson CG. Expression of resistance to tetracyclines in strains of methicillin-resistant *Staphylococcus aureus*. J Antimicrob Chemother. 2000;45(6):763–770. https://doi.org/10.1093/jac/45.6.763

28. Martineau F, Picard FJ, Grenier L, Roy PH, Ouellette M, Bergeron MG. Multiplex PCR assays for the detection of clinically relevant antibiotic resistance genes in staphylococci isolated from patients infected after cardiac surgery. The ESPRIT Trial. J Antimicrob Chemother. 2000;46:527–534. https://doi.org/10.1093/jac/46.4.527

29. Strommenger B, Bartels MD, Kurt K, Layer F, Rohde SM, Boye K, et al. Evolution of methicillin-resistant *Staphylococcus aureus* towards increasing resistance. J Antimicrob Chemother. 2014;69(3):616–622. https://doi.org/10.1093/jac/dkt413

30. Rizzotti L, Simeoni D, Cocconcelli P, Gazzola S, Dellaglio F, Torriani S. Contribution of enterococci to the spread of antibiotic resistance in the production chain of swine meat commodities. J Food Prot. 2005;68(5):955–965. https://doi.org/10.4315/0362-028X-68.5.955

31. Jensen LB, Frimodt-Moller N, Aarestrup FM. Presence of erm gene classes in Gram-positive bacteria of animal and human origin in Denmark. FEMS Microbiol Lett. 1999;170(1):151–158. https://doi.org/10.1111/j.1574-6968.1999.tb13368.x

32. Böckelmann U, Dörries H, Neus Ayuso-gabella M, De Marçay MS, Tandoi V, Levantesi C, et al. Quantitative PCR monitoring of antibiotic resistance genes and bacterial pathogens in three European artificial groundwater recharge systems. Appl Environ Microbiol. 2009;75(1):154–163. https://doi.org/10.1128/AEM.01649-08

33. Brosius J, Palmer M, Kennedy P, Noller H. Complete nucleotide sequence of a 16S ribosomal RNA gene from *Escherichia coli*. Proc Natl Acad Sci USA. 1978;75(10):4801–4805. https://doi.org/10.1073/pnas.75.10.4801

**AUTHORS:**
S'busiso M. Nkosi[1]
Inikile Lupuleza[1]
Siyanda N. Sithole[1]
Zenzile R. Zelda[2]
Anthony N. Matheri[2]

**AFFILIATIONS:**
[1]Technology Station in Chemicals, Mangosuthu University of Technology, Durban, South Africa
[2]Process, Energy and Environmental Technology Station, University of Johannesburg, Johannesburg, South Africa

**CORRESPONDENCE TO:**
S'busiso Nkosi

**EMAIL:**
nkosis@mut.ac.za

**DATES:**
**Received:** 04 Mar. 2021
**Revised:** 25 June 2021
**Accepted:** 28 June 2021
**Published:** 29 Nov. 2021

**HOW TO CITE:**
Nkosi SM, Lupuleza I, Sithole SN, Zelda ZR, Matheri AN. Renewable energy potential of anaerobic mono- and co-digestion of chicken manure, goat manure, potato peels and maize pap in South Africa. S Afr J Sci. 2021;117(11/12), Art. #10362. https://doi.org/10.17159/sajs.2021/10362

**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:**
anaerobic digestion, biomass, biomethane potential, energy mix, environment, renewable energy

**FUNDING:**
Mangosuthu University of Technology, University of Johannesburg, Technology Innovation Agency (South Africa)

# Renewable energy potential of anaerobic mono- and co-digestion of chicken manure, goat manure, potato peels and maize pap in South Africa

The energy sector is an essential part of a country's economy – it drives innovation and advances industrialisation. Coal is the primary source of energy in South Africa. Coal contributes 95% of energy production; coal-fired power also contributes to greenhouse gas emissions, and is thus a hazard to human health and the environment. This calls for an energy mix that is renewable, sustainable, and affordable and that is carbon neutral (climate action). We investigated the potential of anaerobic mono- and co-digestion of goat manure, chicken manure, potato peels, maize pap, and cow manure inoculum for mesophilic recovery of renewable energy using the biomethane potential test. The substrates were characterised through proximate and ultimate analyses to determine the composition preferable for mono- and co-digestion. The key considerations in the determination of both the yield and production rate of methane from digestion of biomass are the substrate composition and characterisation. A high percentage of volatile solids favoured optimum biomethane production as highly volatile components provide microbes with balanced nutrients that enhance metabolic processes to produce biomethane. The mono-digestion process produced lower biomethane than did co-digestion. Higher production of biomethane by co-digestion was due to the balance of the micronutrients and macronutrients that favoured microbial metabolism and regulation of pH.

**Significance:**

- The results highlight the need for appropriate techniques in combining energy and waste management. Biogas could provide solutions for some of South Africa's energy necessities, particularly in rural areas that have abundant biogas substrates in the form of waste from goats and chickens, as well as from kitchen waste.

## Introduction

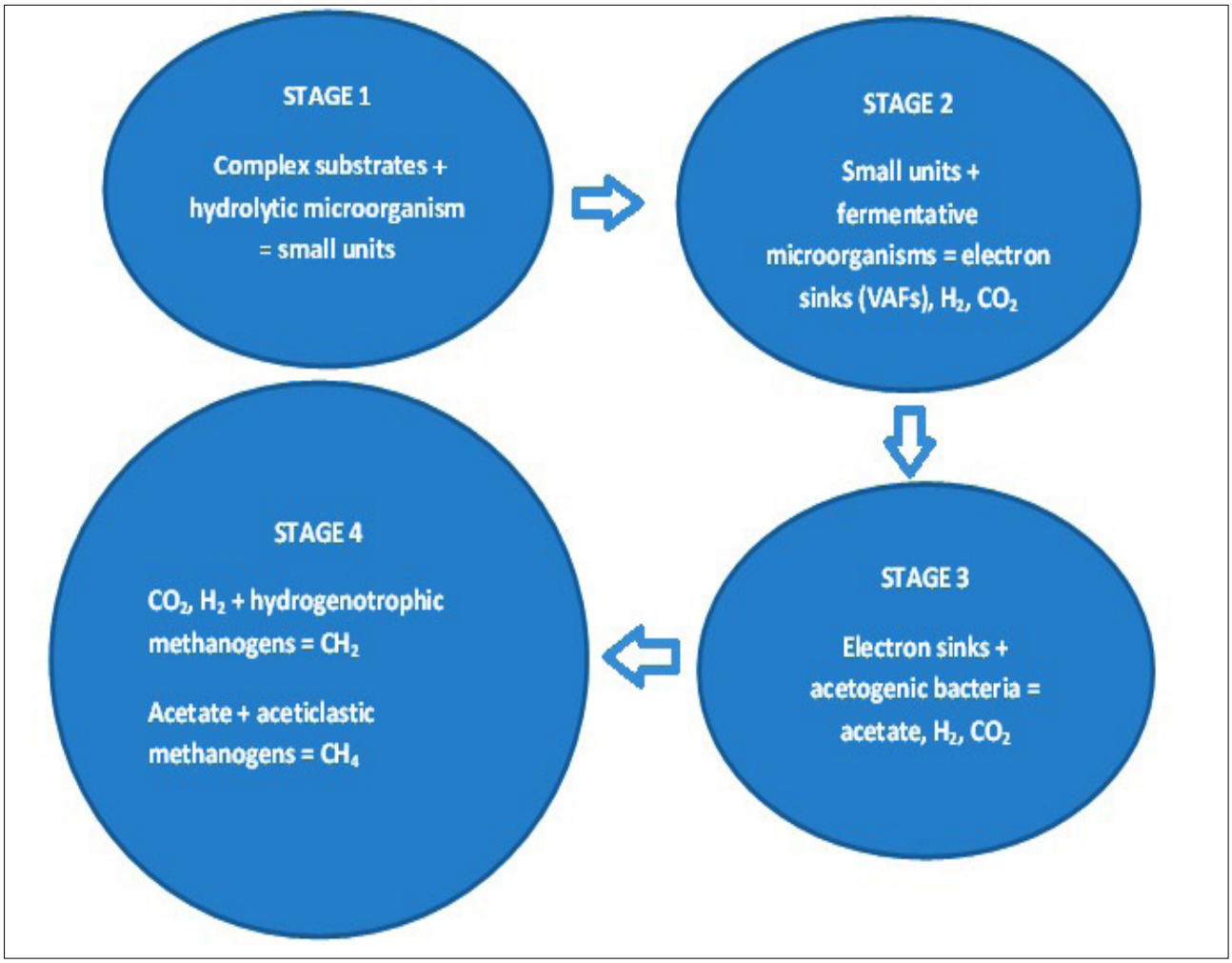
South Africa is fortunate to have an abundance of assorted energy resources that are yet to be exploited and contribute to the energy mix. The primary sources of renewable energy are solar, wind, hydro, and biomass.[1] Renewable energy is taking its rightful place in the South African energy sector and playing a significant role in contributing towards sustainable development.[2] South Africa is ranked as the sixth highest contributor to greenhouse gas emissions worldwide. Carbon dioxide emission per capita is 77% of the total emission in the country. In accordance with the European Union's Renewable Energy Directive[3], energy generation from renewable sources is deemed to be a necessary target for reducing the impact of greenhouse gases, in particular fossil fuel combustion[4]. Negative health impacts and environmental degradation are driving the country's drive towards policies that are in line with Sustainable Development Goals 7 (Affordable and Clean Energy) and 13 (Climate Action).

### *Biomass and anaerobic digestion*

Biomass is currently one of the country's main contributors to renewable energy, with 9–14% of the overall energy mix.[5] Production of energy by anaerobic digestion can be a noble resource channel if appropriately harnessed, as in the case of China and India. Renewable energy provides an easily accessible alternative for rural areas that are off the grid and have decentralised production capabilities. By contrast, anaerobic digestion is a viable technology for the production of biogas derived from organic waste. Anaerobic digestion treatment of organic feedstock to produce biogas offers a two-pronged solution to biomass waste management: generating energy and simultaneously solving the ecological and agrochemical problem. Anaerobic digestion technologies can be divided into three major groups based on the substrate's total solids (TS) content: wet anaerobic digestion with TS less than 15%, dry anaerobic digestion with TS less than 25%, and solid-state anaerobic digestion with TS content of up to 40%.[6] In an anaerobic environment, symbiotic microorganisms convert organic matter into biogas, a constituent mixture of methane ($CH_4$) and carbon dioxide ($CO_2$) together with some nutrients, digestion-resilient organic materials and other cell components like salts. This is a four-stage process: hydrolysis, acid-genesis and acetogenesis are induced by a specific consortium of bacteria, with the final step of methanogenesis undertaken by a consortium of methanogenic archaea (Figure 1).[6]

### *Anaerobic co-digestion*

Additionally, anaerobic co-digestion entails two or more feedstocks being digested simultaneously. This is a sustainable and economically viable option which results in higher yields of methane with added advantages of minimising the challenges associated with mono-digestion. Problems with anaerobic digestion, such as the presence of hazardous materials, unbalanced nutrients or obstinate compounds in the substrate, have rendered anaerobic co-digestion of multi-substrate a common field of research in the advancement and upgrade of conventional anaerobic digestion technology. Work on anaerobic co-digestion has grown dramatically over the last 15 years, demonstrating its capacity to make progress in the production of biogas.[7-10]




*VAF, volatile fatty acids*

**Figure 1:** A schematic of the microbial processes involved in anaerobic digestion.

The main objective of anaerobic co-digestion is to increase methane for renewable energies. As shown in Figure 2, a variety of substrates can be co-digested by blending into different ratios and maintaining optimum conditions needed for the metabolic activities and improvement in the biogas production.

Unlike wastewater, the solid feedstock contains high levels of insoluble organic matter and chemical oxygen demand. These organic matters can be recycled to provide bioenergy. Animal manure is an organic product that is used in horticulture and agriculture as a natural fertiliser. It is a combination of faeces, urine and products that may be used in waste control, such as grass, sand, washing water and other bedding products.[11] Animal manure from livestock farming has been customarily reused as organic fertiliser for the provision of nitrogen and phosphorus for crops and plants. However, the environment was negatively impacted due to increased concentrations of carbon and discharge of soluble phosphorus and nitrogen. Animal manure from dairy cattle[12] and buffalo farming[13] produces large quantities of greenhouse gases, second only to enteric methane. Furthermore, manure generates volatilised ammonia that is up to 70% of the excreted residual nitrogen which enters water systems and natural ecosystems or contributes to climate change and eutrophication through conversion into $N_2O$ emissions.[14]

Goat manure is commonly found in South Africa. It is an excellent substrate for anaerobic digestion due to its optimum range of C/N ratio and robustness of growth. The amount of manure produced by a goat with a typical body weight of 20–40 kg is approximately 0.32–0.63 kg/day, which is comparable to approximately 0.3 tons/year.[15] For instance, a nation with around 1 million goats is estimated to produce approximately 0.3 million tons of organic fertiliser per year.

Chicken manure is a semi-solid biodegradable material. Therefore, it can be used to generate cheap energy.[16] Fresh chicken manure contains 0.4–0.5% phosphorus, 0.8% potassium and 0.9–1.5% nitrogen. It has been reported that the daily production of chicken manure varies from 80 g to 125 g (wet), of which 20–25% of the production comprises total nitrogen-rich solids and 55–65% contains volatile solids (VS) which are a valuable energy source.[17] However, the higher content of nitrogen in chicken manure compared with that in manure from other farm animals (i.e. cattle, pig, horse, goat, sheep, and rabbit)[18-26] makes it unsuitable for the anaerobic digestion process[27]. Ammonia inhibition is one of the common problems with the anaerobic digestion process when using substrates like poultry manure[28] and pig manure[29]. The inhibitory impacts of free ammonia on the methanogen's digestive mechanism was observed by Zhang et al.[30] In general, one of the tricks to avoid ammonia inhibition is to dilute the chicken manure with water[31], which assists in decreasing the high percentage of TS. This dilution step is said to reduce the production of biogas per unit of fermenter volume, increase the use of water and increase the processing costs for the slurry discharge. Several researchers have assessed the co-digestion of chicken manure with other animal manures such as cow manure[32] or pig manure[33], or with other substrates, such as potato peels[34].

Leftover, lost or uneaten food is food waste and can originate throughout the entire food supply chain. This value chain starts from preparation to production, processing, delivery, storage and sale to cooking and presentation.[35] Consumer stage food waste from restaurants, homes, schools, and hospital cafeterias is usually targeted.[36]

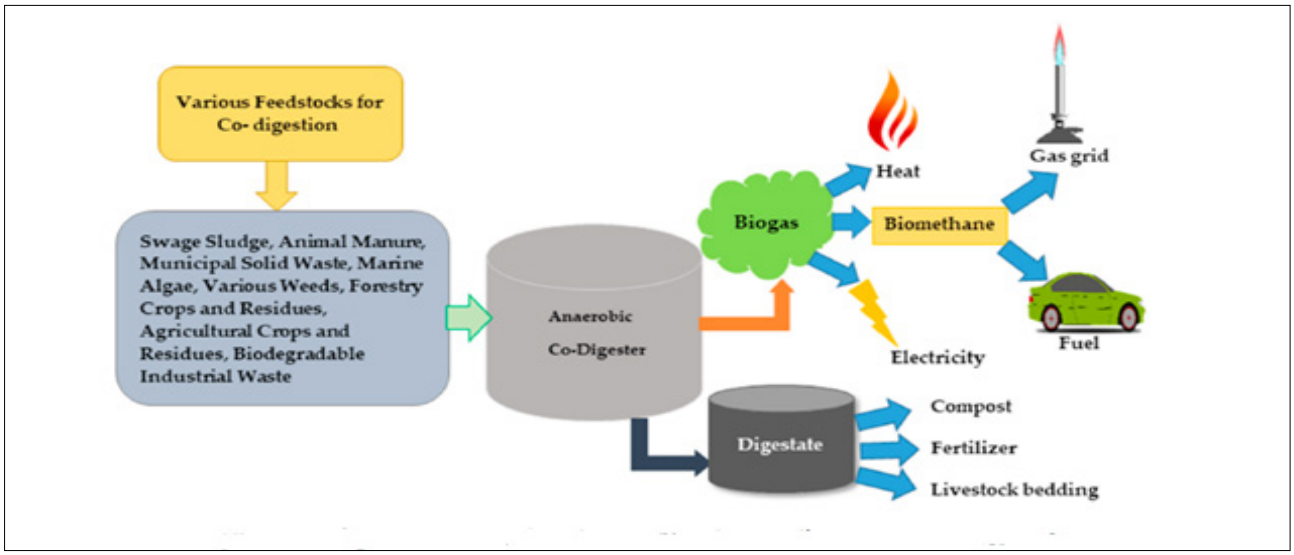



**Figure 2:** A diagram of multi-feedstock co-digestion in energy recovery.

Developing countries generally have scant information with regard to food wastage. Food wastage is globally reported to be higher in developed than in developing countries.[37,38] An annual per capita average of food waste in developed countries ranges from 100 kg to 170 kg, which is double that of developing countries.[39] A few developing countries, including India and China, are considered to have greater challenges in food waste disposal despite the broad total population.[39] More than 40% of the food (about 222 million tons) is lost or wasted in the retail pathway and consumer stages in developed countries, which is nearly as high as the total net food production in sub-Saharan Africa (230 million tons).[39] Although anaerobic digestion technology is a well-known technology for energy recovery, there are still unaddressed issues related to economic and environmental concerns, food security, technology choice, organic loading rates, nature of substrates, parameters variation in different locations globally, disposal of digestate after digestion, and mitigation to climate change.[40,41] In the present study, we investigated the potential of anaerobic mono- and co-digestion of potato peels, maize pap, chicken manure, and goat manure for renewable energy recovery.

## Methodology

### *Substrate quantification*

Food waste was collected from the cafeteria and animal manure samples were collected from the Department of Agriculture farm of the Mangosuthu University of Technology (KwaZulu-Natal Province, South Africa). Samples were collected into 5-L plastic containers, closed airtight, and stored at 4 °C in the refrigerator. The samples were labelled as chicken manure, goat manure, potato peels, maize pap, and inoculum from the cow manure. To pre-treat the samples and minimise the maturing effect, as well as to limit the microbial activities, the feedstock was dried in the oven at 60 °C, reduced to a 2-mm size diameter, and stored in the fridge at 4 °C until utilised. The inoculum was prepared by digesting cow dung samples until no biogas was produced. This was done in order to ensure that the microorganisms digest all the substrate in cow dung completely before adding new substrate to ensure that biogas produced was from new substrate and not from substrate in the cow dung. The retention time for the inoculum preparation was 14 days.

### *Experimental procedure*

The biomethane production rate was determined by feeding the substrates and the inoculum into a batch digester. Chicken manure, goat manure, potato peels, maize pap, and inoculum of cow manure were fed as a mono-substrate for mono-digestion. Chicken manure and potato peels, chicken manure and pap, goat manure and potato peels, goat manure and maize pap, maize pap and potato peels, and a mix of chicken manure, potato peels, goat manure and maize pap were co-digested in the ratio of 1:1. The substrates were fed with the control. The conditions were set at mesophilic temperatures of 37 °C and the working pH of 6.5–7.5 was adjusted using sodium hydroxide and sulfuric acid. A bioprocess controller (AMPTS II) by Dürr Systems, Inc. (De Pere, WI, USA) was used to perform the biomethane potential test. The AMPTS II consisted of an automated digester, $CO_2$ fixing unit, and biomethane collection unit. Bioreactors with a volume of 500 mL and a headspace of 100 mL were used. Sodium hydroxide was used to remove $CO_2$ from the production of biogas to biomethane. A 3-M sodium hydroxide solution was used to remove $CO_2$ and $H_2S$. A pH indicator solution with 0.4% thymolphthalein was applied to the sodium hydroxide solution where the pH indicator was used as a scrubber. Before preparing the substrates and feeding into the digesters, the prepared NaOH with pH indicator was used to determine the saturation point for the cleaning solution to be replaced. Thus, the digesters were purged with nitrogen gas to create an anaerobic state by discharging the oxygen. The gas that exited the $CO_2$ fixing unit was sent to the stream cell (gas collection unit) and assessed on a daily basis using water downward displacement technique until the retention time was completed. Figure 3 shows the bioprocess controller (AMPTS II).

### *Analytical techniques*

#### Proximate analysis

The chemical and physical compositions, moisture content, total solids (TS), volatile solids (VS), and hydrogen potential (pH) were calculated using standard methods (APHA 1995).[42]

#### Ultimate analysis

A Flash 2000 CHNS-O element analyser (Thermo Fisher Scientific Inc., Waltham, MA, USA) fitted with an autosampler was used to analyse the elemental (C, H, N, S) composition of the substrate. In replicates per sample, a dry mass of 1 mg of each substratum was weighed into a tin capsule. CHNS was then determined by an autosampler by placing the samples in an electrolyte-filled quartz reactor and then inserting them into the reactor cell. After combustion in an oxygen-rich environment, the gases given off were carried by a helium flow past a copper-filled layer, through a gas chromatography column where the combustion gases were separated and detected by a detector (Thermal Conductivity Detector, Waltham, MA, USA) with a column oven temperature of 65 °C detection.


**Figure 3:** The bioprocess controller (AMPTS II) used to perform the biomethane potential test. The AMPTS II consisted of an automated digester, carbon dioxide fixing unit, and biomethane collection unit.

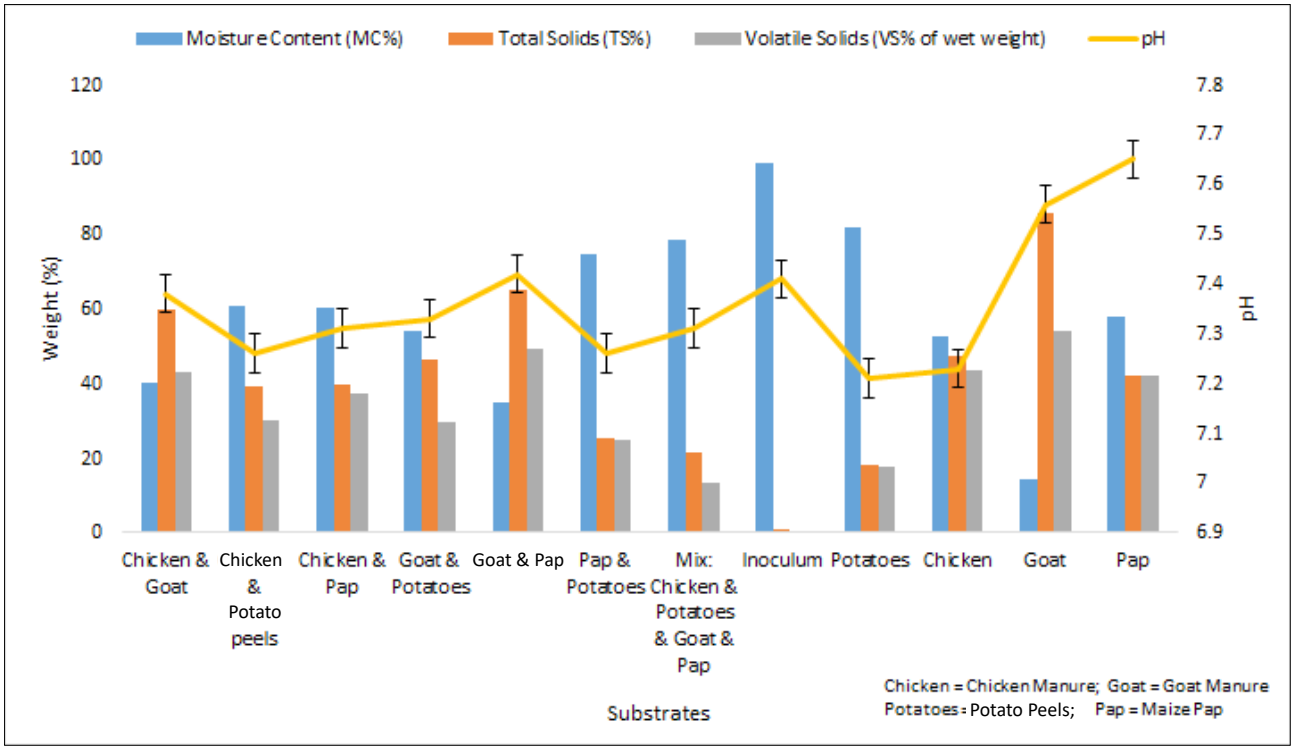



**Figure 4:** Results of the proximate analysis of the substrates and absolute pH on biogas accumulation.

## Results and discussion

### *Characterisation of substrate*

#### Proximate analysis

Characteristics of the mono-substrates chicken manure (CM), goat manure (GM), potato peels (PP), maize pap (MP) and inoculum co-substrates CM + GM, CM + MP, GM + MP, CM + PP, GM + PP, MP + PP are shown in Figure 4. Substrate composition and characterisation can be a major factor in deciding the levels of methane production and methane generation from biomass digestion. VS are the organic component of TS that biodegrade, where TS is the total amount of suspended solids and dissolved solids. Optimum production of biomethane is attained when the %VS is low, as this helps bacteria to provide volatile compounds with methane and metabolic processes. VS therefore play an important role in optimal digestion, because they allow the digestion process to generate high-quality biomethane.

Goat manure and chicken manure had a %VS of 43.46% and 54.28%, respectively. These values indicate a good potential to convert biosolid to bioenergy. The inoculum had lower VS due to initial digestion during the preparation. Potato peels had the highest moisture content in the composition. Due to variation in composition, this called for co-digestion of samples to reach the optimum range for the digestion of substrate to energy (biomethane). The VS/TS ratios for mono-substrates were 63.52%, 91.45%, 99.76% and 96.61% for GM, CM, MP and PP, respectively. The higher the VS/TS ratio, the higher the organic content, which is favourable for the production of biomethane.

### Ultimate analysis

The C/N ratio was imperative in the microscopic organism's stability within the anaerobic process. However, examination of the organic components showed that CM, PP, and MP had C/N ratios of 42.91%, 99.70% and 37.30%, respectively, which are unsuitable for anaerobic digestion and out of range of the optimum (10 to 30) C/N ratio.[42,43] However, GM had a C/N ratio of 20.47, which is within the optimum range. A high C/N ratio leads to a deficiency within the anaerobic digestion system due to the methanogens' fast nitrogen consumption leading to lower gas production. A lower C/N ratio indicates high nitrogen that can transform into ammonia. Ammonia inhibits the microbes' activities. Co-digestion enhanced the nutrient balance of the substrates and provided an optimal C/N range. Figure 5 shows the results of the ultimate analysis of the mono-substrates.

## *Biomethane potential test of biomass*

### Mono-digestion of biomass

The biomethane potential test was vital in assessing the biomethane production of the respective organic feedstock amid its anaerobic deterioration. Figure 6 shows the biomethane production from the mono-digestion of substrates.

The retention time for substrate digestion to yield biomethane was 21 days. The organic loading rate ratio of substrate to inoculum was 1:2 of %VS. It was observed that the inoculum produced lower methane than did the other substrates because it was previously digested for 14 days and the nutrients were exhausted. Maize pap produced the most methane (1650.8 NmL $CH_4$/g VS) because of the high nutrient composition and balance. Carbohydrates are generally considered to be rapidly degradable. The shortest lag phase as observed in Figure 6 was due to the introduction of inoculum that supplied microorganisms. Within 8 days, the production of biomethane was at equilibrium due to the rapid aggregation of unstable greasy acids that led to a reduction in biomethane production. Lower production of the biomethane was attributed to lower growth and metabolism of the methanogens due to poor nutritional structure (C/N ratio) and insufficient micronutrients. PP, GM, and CM produced 1423.6, 726.9, and 120.7 NmL $CH_4$/g VS, respectively.

### Co-digestion of biomass

Figure 7 shows the biomethane production from the co-digestion of substrates. It was observed that most of the highest biomethane accumulation was achieved within the initial 9 days of retention time for all the co-digested substrates except for MP and PP. After the ninth day, all production became steady and reached equilibrium for the remainder of the 21 experimental days. The highest biomethane production of 1332.2 NmL $CH_4$/g VS was achieved for GD and PP substrates. This was due to the availability and balance of nutrients, and microbial balance. The CM and MP substrates recorded the lowest biomethane potential of 474.6 NmL $CH_4$/g VS. In the anaerobic digestion process, a synergistically decomposition of organic matter caused a bacterial consortium generating biogas and biomethane. Inhibition in some substrates occurred because of imbalanced rates between hydrolysis and methanogenesis. It was necessary to have a suitable balance between those levels to obtain higher biomethane production. A rapid pathway for methanogenesis was necessary to prevent the over accumulation of the organic acid and reduction in pH, as well as high ammonia and an increase in pH to a degree that would inhibit methanogenesis. For the production of methane-forming microbes, certain nutrient components are required.[44] Relevant trace elements such as nickel, cobalt, iron and molybdenum are important for the ideal development and production of biomethane because they stimulate methanogenic activity.

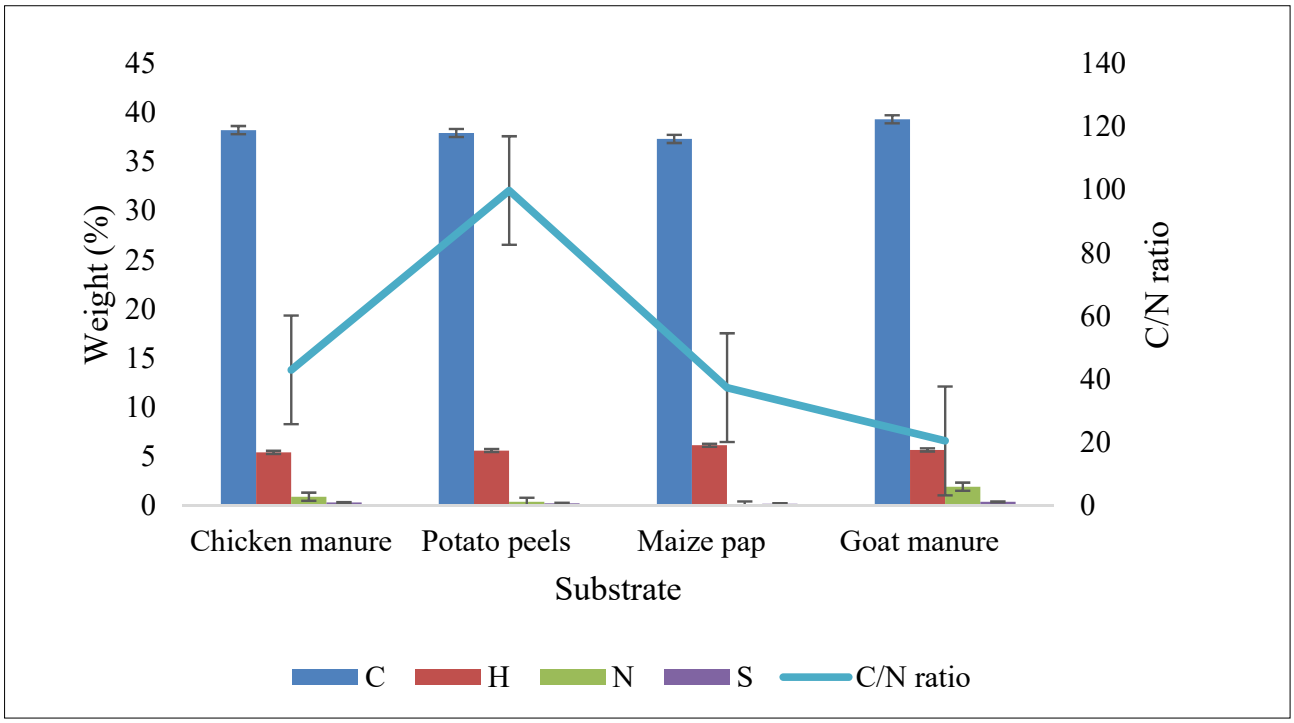



**Figure 5:** Results of the ultimate analysis of the substrates.

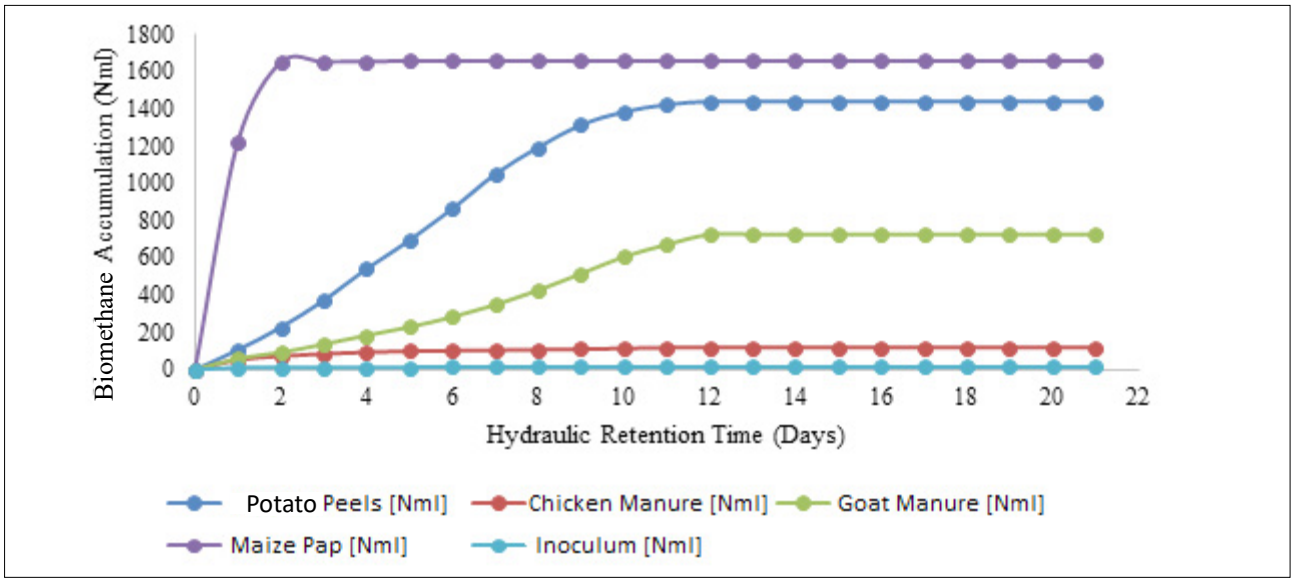



**Figure 6:** Biomethane production through mono-digestion of substrates.

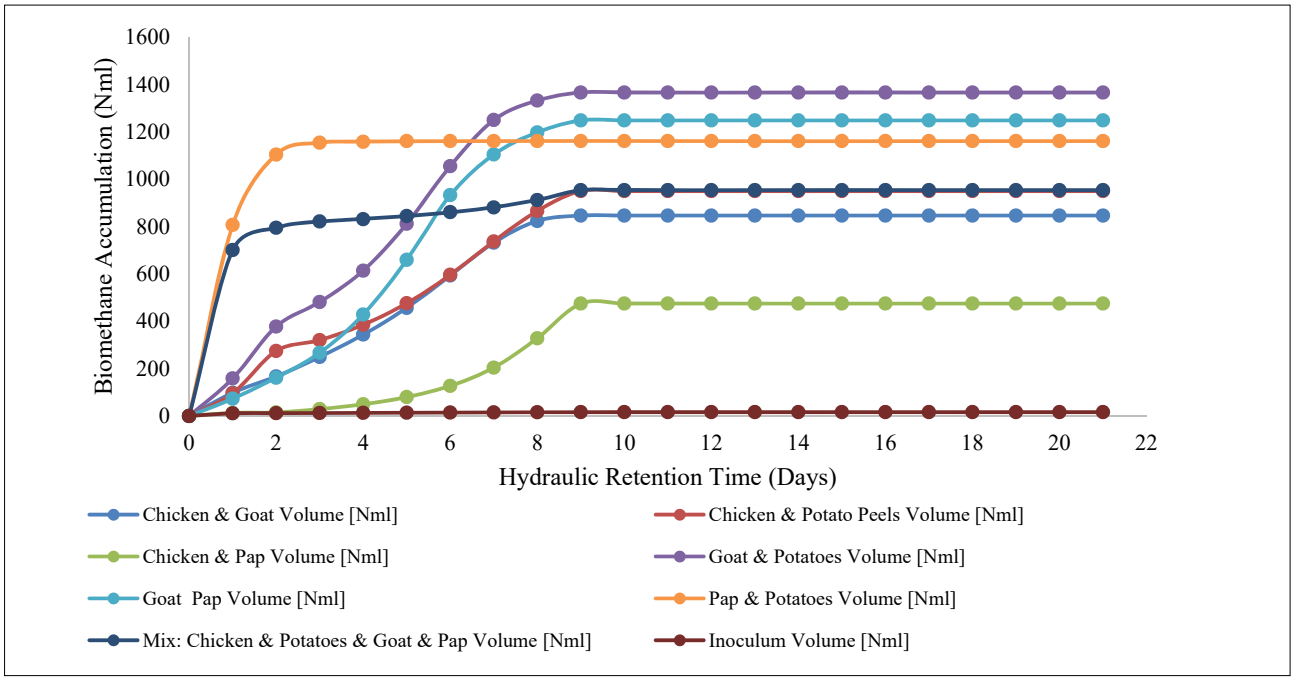



**Figure 7:** Biomethane production through co-digestion of substrates.

The microbial activity in biomethane development was stimulated by boron, manganese, molybdenum, aluminium and selenium in substrates.[45-47] According to Duran and Speece[48], the addition of metal ions to anaerobic digesters plays a key role in improving the performance of the anaerobic digestion system through co-digestion. Co-digestion of substrates enhanced nutrients to microbial balance.

The influence of pH on biomethane production is shown in Figure 4. The highest biomethane production of 1332.2 NmL $CH_4$/g VS was observed at a pH of 7.11, whereas productions of 1248.2, 1154.2, 949.2, 846.4 and 474.6 NmL $CH_4$/g VS were observed at a pH of 5.97, 7.15, 7.52 and 7.54, respectively.

Chicken manure and potato peel substrates had lower pH of 7.23 and 7.21, respectively, within the range of 6.5–7.5. Goat manure and maize pap substrates had higher pH of 7.56 and 7.65, respectively, slightly above the optimum range. Co-digestion of substrates assisted in adjusting the pH to 7.31, within the optimum range.[48-51]

## Conclusions

Volatile solids, moisture content and total solids played an important role in optimal digestion, by enabling the digestion process to generate high-quality biomethane. Goat manure and chicken manure had %VS of 43.46% and 54.28%, indicating a good potential to convert biosolids to bioenergy. Higher VS/TS ratios of the organic content enhanced higher production of biomethane. Co-digestion regulated nutrients and created a nutrient balance of the substrates that were out of the optimum range (C/N ratio of 10–30). Mono-digestion of maize pap resulted in the highest biomethane production of 1650.8 NmL $CH_4$/g VS because of the high nutrient composition and balance. Co-digestion of goat manure and potato peels resulted in the highest biomethane production of 1332.2 NmL $CH_4$/g VS among the co-digested substrates. This was due to the

nutrient and microbial balance. The anaerobic digestion process assisted in energy recovery. Co-digestion addressed issues related to economic and environmental concerns, microbial-nutrient balance, organic loading rates, regulation of parameter variation, and mitigation to climate change. Co-digestion is efficient, economically viable, produces higher yields of methane and has the ability to mitigate some of the problems that may arise from mono-digestion. The problems in anaerobic digestion – such as unbalanced nutrients, the presence of poisonous materials or persistent elements in the substrates – make co-digestion a viable technology.

## Acknowledgements

We acknowledge the Technology Station in Chemicals, Mangosuthu University of Technology (Durban, South Africa), Process, Energy & Environmental Technology Station, University of Johannesburg (Johannesburg, South Africa), and the Technology Innovation Agency for funding and knowledge transfer and Ms Xolile Veronica Ngubane and Ms Esther Nabadda for proofreading and editing the manuscript.

## Competing interests

We have no competing interests to declare.

## Authors' contributions

S.M.N.: Conceptualisation, methodology, sample analysis, project management, student supervision, writing – the initial draft, writing – revisions, funding acquisition. I.L.: Data collection, sample analysis. S.N.S.: Data collection, sample analysis. Z.R.Z.: Project management, project leadership. A.N.M.: Methodology, writing – the initial draft, writing – revisions.

## References

1. Omar A, Haitham A, Frede B. Renewable energy resources: Current status, future prospects, and their enabling technology. Renew Sustain Energy Rev. 2014;39:748–764. https://doi.org/10.1016/j.rser.2014.07.113
2. Pegels A. Renewable energy in South Africa: Potentials, barriers, and options for support. Energy Policy. 2010;38:4945–4954. https://doi.org/10.1016/j.enpol.2010.03.077
3. Directive 2009/28/EC of the European parliament and of the Council of 23 April 2009 on the promotion of the use of energy from renewable sources and amending and subsequently repealing directives 2001/77/EC and 2003/30/EC. Brussels: European Commission; 2009.
4. Olivier JGJ, Schure KM, Peters AW. Trends in global $CO_2$ and total greenhouse gas emissions. PBL publication number 2674. The Hague: PBL Netherlands Environmental Assessment Agency; 2017.
5. Banks D, Jason S. The potential contribution of renewable energy in South Africa. Johannesburg: Sustainable Energy and Climate Change Project of Earthlife Africa; 2006.
6. Mamun A, Torii S. Anaerobic co-digestion technology in solid wastes treatment for biomethane generation. Int J Sustain Energy. 2015;36:462–472. https://doi.org/10.1080/14786451.2015.1043302
7. Kwietniewska E, Tys J. Process characteristics, inhibition factors, and methane yields of the anaerobic digestion process, with particular focus on microalgal biomass fermentation. Renew Sustain Energy Rev. 2014;34:491–500. https://doi.org/10.1016/j.rser.2014.03.041
8. Gelegenis J, Georgakakis D, Angelidaki I, Christopoulou N, Goumenaki M. Optimization of biogas production from olive-oil mill wastewater, by co-digesting with diluted poultry-manure. Appl Energy. 2007;84:646–663. https://doi.org/10.1016/j.apenergy.2006.12.001
9. Karagiannidis A, Perkoulidis G. A multi-criteria ranking of different technologies for the anaerobic digestion for energy recovery of the organic fraction of municipal solid wastes. Bioresour Technol. 2009;100:2355–2360. https://doi.org/10.1016/j.biortech.2008.11.033
10. Hagos K, Zong J, Li D, Liu C, Lu X. Anaerobic co-digestion process for biogas production: Progress, challenges, and perspectives. Renew Sustain Energy Rev. 2017;76:1485–1496. https://doi.org/10.1016/j.rser.2016.11.184
11. Rao JR, Watabe M, Stewart TA, Millar BC, Moore JE. Pelleted organo-mineral fertilizers from composted pig slurry solids, animal wastes and spent mushroom compost for amenity grassland. J Waste Manag. 2007;27:1117–1128. https://doi.org/10.1016/j.wasman.2006.06.010
12. Aguirre-Villegas HA, Larson R. Evaluating greenhouse gas emissions from dairy manure management practices using survey data and lifecycle tools. J Clean Prod. 2017;143:169–179. https://doi.org/10.1016/j.jclepro.2016.12.133
13. Sabia E, Napolitano F, Claps S, De Rosa G, Braghieri A, Pacelli C. Dairy buffalo life cycle assessment as affected by heifer rearing system. J Clean Prod. 2018;192:647–655. https://doi.org/10.1016/j.jclepro.2018.04.158
14. Hristov AN, Zaman S, Vander Pol M, Ndegwa P, Campbell L, Silva S. Nitrogen losses from dairy manure estimated through nitrogen mass balance and chemical markers. J Environ Qual. 2001;38:2438–2448. https://doi.org/10.2134/jeq2009.0057
15. Yadav SK, Subhash B, Yadav MK, Singh K, Yadav GS, Pal S. A review of organic farming for sustainable agriculture in northern India. Int J Agron. 2013;2013, Art. #718145. https://doi.org/10.1155/2013/718145
16. Abouelenien F, Fujiwara W, Namba Y, Kosseva M, Nishio N, Nakashimada Y. Improved methane fermentation of chicken manure via ammonia removal by biogas recycle. Bioresour Technol. 2010;101:6368–6373. https://doi.org/10.1016/j.biortech.2010.03.071
17. Moral R, Moreno-Caselles J, Perez-Murcia M, Perez-Espinosa A, Rufete B, Paredes C. Characterisation of the organic matter pool in manures. Bioresour Technol. 2005;96:153–158. https://doi.org/10.1016/j.biortech.2004.05.003
18. Olowoyeye J. Comparative studies on biogas production using six different animal dungs. J Biol Agric Healthcare. 2013;3:7–12.
19. Triolo JL, Ward AJ, Pedersen L, Sommer SG. Characteristics of animal slurry as key biomass for biogas production in Denmark. In: Matovic MD, editor. Biomass now – sustainable growth and use. London: InTech; 2013.
20. Chen F, Yu G, Li W, Liu FW, Zhang WP, Bu YS, et al. Maximal methane potential of different animal manures collected in the northwest region of China. Int J Agric Biol. 2017;10:202–208.
21. Buratti FC. Biogas production from different substrates in an experimental continuously stirred tank reactor anaerobic digester. Bioresour Technol. 2009;100:5783–5789. https://doi.org/10.1016/j.biortech.2009.06.013
22. Pham CH, Triolo JM, Cu TTT, Pedersen L, Sommer SG. Validation and recommendation of methods to measure biogas production potential of animal manure. Asian Australas J Anim Sci. 2013;26:864–873. https://doi.org/10.5713/ajas.2012.12623
23. Andrade WR, Xavier CAN, Coca FG, Arruda LDO, Santos TMB. Biogas production from ruminant and monogastric animal manure co-digested with manipueira. Arch de Zootec. 2016;65:251–380. https://doi.org/10.21071/az.v65i251.699
24. Kafle GK, Chen L. Comparison on batch anaerobic digestion of five different livestock manures and prediction of biochemical methane potential (BMP) using different statistical models. J Waste Manag. 2016;48:492–502. https://doi.org/10.1016/j.wasman.2015.10.021
25. Budiyono B, Widiasa IN, Johari S, Sunarso S. Increasing biogas production rate from cattle manure using rumen fluid as inoculums Int J Eng Sci. 2014;6:31–38. https://doi.org/10.12777/ijse.6.1.31-38
26. Osman GAM, Elhasan HE, Hassan AB. Effect of cow rumen fluid concentration on biogas production from goat manure. Sudan J Agric Sci. 2015;2:1–7.
27. Bujoczek G, Oleszkiewicz J, Sparling R, Cenkowski, S. High solid anaerobic digestion of chicken manure. J Agric Eng Res. 2000;76:51–60. https://doi.org/10.1006/jaer.2000.0529
28. Gangagni R, Sasi Kanth A, Reddy T, Surya Prakash S, Vanajakshi J, Joseph J, et al. Biomethanation of poultry litter leachate in UASB reactor coupled with ammonia stripper for enhancement of overall performance. Bioresour Technol. 2008;99:8679–8684. https://doi.org/10.1016/j.biortech.2008.04.016
29. Hansen KH, Angelidaki I, Ahring BK. Anaerobic digestion of swine manure: Inhibition by ammonia. Water Res. 1998;32:5–12. https://doi.org/10.1016/S0043-1354(97)00201-7
30. Zhang C, Yuan Q, Lu Y. Inhibitory effects of ammonia on methanogen more transcripts in anaerobic digester sludge. FEMS Microbiol Ecol. 2014;87:368–377. https://doi.org/10.1111/1574-6941.12229

31. Niu Q, Qiao W, Qiang H, Li YY. Microbial community shifts and biogas conversion computation during steady, inhibited, and recovered stages of thermophilic methane fermentation on chicken manure with a wide variation of ammonia. Bioresour Technol. 2013;146:223–233. https://doi.org/10.1016/j.biortech.2013.07.038

32. Wang X, Yang G, Feng Y, Ren G, Han X. Optimizing feeding composition and carbon-nitrogen ratios for improved methane yield during anaerobic co-digestion of dairy, chicken manure, and wheat straw. Bioresour Technol. 2012;120:78–83. https://doi.org/10.1016/j.biortech.2012.06.058

33. Borowski S, Domanski J, Weatherley L. Anaerobic co-digestion of swine and poultry manure with municipal sewage sludge. J Waste Manag. 2014;34:513–521. https://doi.org/10.1016/j.wasman.2013.10.022

34. Ali S, Shah TA, Afzal A, Tabbassum R, Evaluating the co-digestion effects on chicken manure and rotten potatoes in batch experiments. Int J Biosci. 2017;10:150–159. https://doi.org/10.12692/ijb/10.6.150-159

35. Lin CSK, Pfaltzgraff ALA, Herrero-Davila L, Mubofu EB, Abderrahim S, Clark JH, et al. Food waste as a valuable resource for the production of chemicals, materials, and fuels. Current situation and global perspective. Energy Environ Sci. 2013;6:426–464. https://doi.org/10.1039/c2ee23440h

36. Parfitt J, Barthel M, Macnaughton S. Food waste within food supply chains: Quantification and potential for change to 2050. Philos Trans R Soc B. 2010;365:3065–3081. https://doi.org/10.1098/rstb.2010.0126

37. United Nations. Responsible consumption and production: Why it matters [document on the Internet]. c2019 [cited 2020 May 20]. Available from: https://www.un.org/sustainabledevelopment/wp-content/uploads/2019/07/12_Why-It-Matters-2020.pdf

38. Dunga TN, Sena B, Chen C, Kumare G, Lina C. Food waste to bioenergy via anaerobic processes. Energy Procedia. 2014;61:307–312. https://doi.org/10.1016/j.egypro.2014.11.1113

39. Gustafson J, Cederberg C, Sonesson U, Van Otterdijk R, Meybeck A. Global food losses and food waste: Extent, causes, and prevention. Rome: Food and Agriculture Organization of the United Nations; 2011.

40. Food and Agriculture Organization of the United Nations (FAO). Global food losses and food waste [document on the Internet]. c2011 [cited 2020 May 20]. Available from: http://www.fao.org/docrep/014/mb060e/mb060e00.pdf

41. Nishio N, Nakashimada Y. Recent development of anaerobic digestion processes for energy recovery from wastes. J Biosci Bioeng. 2007;103:105–112. https://doi.org/10.1263/jbb.103.105

42. Rice EW. Standard methods for the examination of water and wastewater. Vol. 10. Washington DC: American Public Health Association; 2012:

43. Gunaseelan VN. Biochemical methane potential of fruits and vegetable solid waste feedstocks. Biomass Bioenergy. 2004;26:389–399. https://doi.org/10.1016/j.biombioe.2003.08.006

44. Romano TT, Zhang RH. Co-digestion of onion juice and wastewater sludge using an anaerobic mixed biofilm reactor. Bioresour Technol. 2008;99:631–637. https://doi.org/10.1016/j.biortech.2006.12.043

45. Montingelli ME, Tedesco S, Olabi AG. Biogas production from algal biomass: A review. Renew Sustain Energ Rev. 2015;43:961–972.

46. Karlsson A, Truong XB, Gustavsson J, Svensson BH, Nilsson F, Ejlertsson J. Anaerobic treatment of activated sludge from Swedish pulp and paper mills – biogas production potential and limitations. Environ Technol. 2012;32:1559–1571.

47. Van Dijk L, Roncken G, Combinatie van biologische afvalwaterzuivering en membraan-technologie [Combination of biological wastewater treatment and membrane technology]. Water J. 1994:41–48. Dutch.

48. Duran M, Speece R. Biodegradability of residual organics in the effluent of anaerobic processes. Environ Technol. 1999;20:597–605. https://doi.org/10.1080/09593332008616854

49. Matheri AN, Sethunya VL, Belaid M, Muzenda E. Analysis of the biogas productivity from dry anaerobic digestion of organic fraction of municipal solid waste. Renew Sustain Energy Rev. 2018;81:2328–2334. https://doi.org/10.1016/j.rser.2017.06.041

50. Matheri AN, Ntuli F, Ngila JC, Seodigeng T, Zvinowanda C, Njenga CK. Quantitative characterization of carbonaceous and lignocellulosic biomass for anaerobic digestion. Renew Sustain Energy Rev. 2018;92:9–16. https://doi.org/10.1016/j.rser.2018.04.070

51. Matheri AN, Mbohwa C, Ntuli F, Belaid M, Seodigeng T, Ngila JC, et al. Waste to energy bio-digester selection and design model for the organic fraction of municipal solid waste. Renew Sustain Energy Rev. 2018;82:1113–1121. https://doi.org/10.1016/j.rser.2017.09.051

**AUTHORS:**
Eki T. Aisien[1]
Felix A. Aisien[2]

**AFFILIATIONS:**
[1]Department of Environmental Management and Toxicology, University of Benin, Benin, Nigeria
[2]Department of Chemical Engineering, University of Benin, Benin, Nigeria

**CORRESPONDENCE TO:**
Felix Aisien

**EMAIL:**
aibue.aisien@uniben.edu

**DATES:**
**Received:** 24 Mar. 2021
**Revised:** 30 June 2021
**Accepted:** 08 July 2021
**Published:** 29 Nov. 2021

**HOW TO CITE:**
Aisien ET, Aisien FA. Photodegradation and kinetics of edible oil refinery wastewater using titanium dioxide. S Afr J Sci. 2021;117(11/12), Art. #10483. https://doi.org/10.17159/sajs.2021/10483

**ARTICLE INCLUDES:**
☒ Peer review
☐ Supplementary material

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

**EDITOR:**
Teresa Coutinho

**KEYWORDS:**
edible oil wastewater, photocatalytic treatment, physicochemical parameters, titanium dioxide, water quality standards

**FUNDING:**
None

# Photodegradation and kinetics of edible oil refinery wastewater using titanium dioxide

Edible oil refinery wastewater (EORW) is one source of environmental pollution in Nigeria. The treatment of EORW before discharge into the environment remains a significant challenge in the edible oil refinery industries. This research was aimed at photocatalytic treatment of EORW using a batch photocatalytic reactor with titanium dioxide photocatalyst. We investigated the physicochemical parameters: chemical oxygen demand (COD), biological oxygen demand ($BOD_5$), oil and grease, phenol, chloride ($Cl^-$), total suspended solids, sulfate ($SO_4^{2-}$), and phosphate ($PO_4^{3-}$) using American Public Health Association methods. The results showed that the reduction efficiency of the treated EORW with $TiO_2$ catalyst ranged between 65.8% ($PO_4^{3-}$) and 87.0% (COD), and the improvement in efficiency was 54.1% (pH) and 60.8% dissolved oxygen. However, the results showed no significant difference ($p<0.05$) in the control treatment without catalyst. The biodegradability of EORW increased from 0.196 to 0.32. It was observed that the optimum values were an initial EORW concentration of 100 mL/L, irradiation time of 90min, catalyst dose of 1.25 g/L, and an agitation speed of 900 rpm. The kinetics of the photodegradation process was well described by the pseudo-first-order equation ($R^2>0.96$) and pseudo-second-order equation ($R^2>0.98$). The intra-particle diffusion model fairly represented the diffusion mechanism with an $R^2$ value of 0.806. The treated EORW met the most acceptable water quality standards for discharged effluent according to the maximum permissible limits of the Nigerian National Environmental Standards and Regulations Enforcement Agency.

**Significance:**

- Photocatalytic treatment of EORW gave negligible results in the absence of a photocatalyst.
- The photocatalytic degradation of EORW improved its biodegradability.
- Photocatalytic treatment of EORW significantly reduced the pollutants in the wastewater.
- The pseudo-first-order equation ($R^2>0.96$) and pseudo-second-order equation ($R^2>0.98$) well described the photodegradation process of EORW.

## Introduction

There is an enormous consumption of fresh water and energy in the processing of edible oil. Also, unwanted liquid, solid and gaseous wastes are generated together with the desired products (edible oils).[1] Edible oil refinery wastewater is generated from the degumming, deacidification, and deodorisation and neutralisation processes.[2] The indiscriminate disposal of this waste and wastewater results in severe environmental pollution and health hazards. Thus, there is an urgent need to manage and treat this waste and wastewater to qualitatively and quantitatively reduce its pollutant loads and reuse the waste and wastewater. The essence of treating large amounts of wastewater generated from edible oil refineries is to remove contaminants before discharge into the environment.[3] However, edible oil refinery wastewater (EORW) varies in characteristics and quantity from one industry to another. The composition of EORW from the same industry also differs from day to day.[4]

Some researchers have used an activated sludge reactor[5], electrocoagulation[6-9], coagulation-flocculation[10], electro-oxidation[11], and a combination of advanced oxidation processes and biological[12] methods, for example, in the treatment of EORW, and achieved chemical oxygen demand (COD) removal efficiencies of between 76% and 99%. Also, Ma et al.[13] and Kalat and Yüceer[14] used a biological method to remove 68–99% oil and grease. Esteves et al.[15] reported that Fenton's process gave a 96.9% reduction of phenolic content and biodegradability enhancement. These methods' drawbacks include inadequate removal of dissolved organic matter and nutrients, scarce nitrifying bacteria and low processing load per volume, requirement of expensive equipment, high handling costs, and production of chemical sludge.[16,17] The EORW contains phenols, heavy metals, catalysts, fats and oils, oxidisable substances, and other complicated dissolved recalcitrant compounds[3], hence the need to use a physicochemical method such as photocatalysis for its treatment. Liu et al.[18] and Welz[19] reported that photocatalytic pretreatment of excess sludge resulted in increased biodegradability of recalcitrant matter, allowing for further treatment of wastewater by continuous biological treatment.

The complexity, characteristic variations, and quantity of EORW have led to the investigation of photocatalytic degradation in this study. Other added advantages of photocatalysis include quick and efficient removal and elimination of organic matter and pathogens[20], mineralisation of a wide range of organic compounds into carbon dioxide, water, and inorganic ions, and intense activity of the process[21-23]. Moreover, its ability to increase the contaminants' biodegradability enable it to integrate with biological methods for process cost-effectiveness.[24] The biodegradability (biological oxygen demand (BOD)/COD) of organic matter is an important factor for evaluating both the treatability of organic matter in wastewater treatment plants and biogeochemical roles of effluents in the receiving water.[25] The ratio of BOD/COD has been used as a good surrogate for the biodegradability. Biodegradability determines the sustainability and suitability of wastewater for further treatment.[26] Dhanke and Wagh[27] reported that

increasing the biodegradability index is good for biological degradation of wastewater. Nagar and Devra[28] stated that the use of advanced oxygen processes will increase the biodegradability of textile wastewater by using silver nanoparticles.

The most widely used photocatalyst for the photocatalytic degradation process is titanium dioxide ($TiO_2$). It is low cost, non-toxic, chemically stable, environmentally friendly, and has a high oxidative power.[29,30] However, $TiO_2$ powder is subject to agglomeration that reduces its surface area, hence its catalytic efficiency. Also, there is the difficulty of separation and recovery of the $TiO_2$ catalyst at the end of the photodegradation process. The methods that have been used to address these problems include immobilisation on various supports such as silica, silicon carbide, perlite, fly ash, periwinkle shell ash, zeolites, clay, and activated carbon.[31-36] The immobilisation of $TiO_2$ on solid supports involves different preparation methods such as hydrothermal, sol-gel, precipitation, metal organic chemical vapour deposition, pyrolysis, impregnation, microwave-assisted synthesis, and sonochemical treatment.[37] This immobilisation has proven to be very efficient in removing organic contaminants due to the absorptive nature of the $TiO_2$ photocatalyst surface.[38] The high absorptive efficiency of the $TiO_2$ composite is maintained over a long period.[39]

The problem of contamination of surface and underground water by untreated EORW remains a challenge in Nigeria. Therefore, this study was focused on the efficient removal of pollutants from EORW using the photocatalytic degradation process. We also performed kinetic modelling using pseudo-first-order, pseudo-second-order, and intra-particle diffusion models to determine which of the kinetic model(s) best fit the experimental data. Hence, the research was aimed at investigating the use of photocatalysis for the treatment of EORW and the process kinetics. Our objectives were to evaluate the performance efficiency of $TiO_2$ as a photocatalyst regarding its ability to reduce $BOD_5$ and COD in EORW and optimise the treatment process through kinetic models.

## Materials and methods

### Materials collection

Three 50-L clean and sterile plastic containers were used to collect raw EORW samples from three discharge points at Egharevbe oil mill in Ebuobanosa-Benin, Nigeria (N6°20'1.32" E5°36'0.53"). All chemicals/reagents used, including commercial titanium dioxide ($TiO_2$) [99.5% anatase], were of analytical grade and purchased from Sigma-Aldrich Co. Ltd (Gillingham, UK), and Qualikems Fine Chemical Ltd (Vadodara, India).

### Material preparation

The EORW samples were preserved at 4 °C in a refrigerator before treatment. Before commencing treatment, the EORW samples were removed from the refrigerator and held at room temperature (28±2 °C) for 2 h.

The raw EORW samples were prepared into different initial concentrations by diluting them with deionised water. The initial concentration range was 100 mL/L to 250 mL/L. We prepared the $TiO_2$-EORW solution at different concentrations ranging from 0.5 g/L to 1.5 g/L.

### Experimental instrumentation

We used a spectrophotometer 910 model; multi-meter tester 2010 model, HACH colourimeter 402 model, pH meter 3010 model dissolved oxygen meter HI 981193, BOD/COD/total organic carbon meter (Aquadax), and total dissolved solids meter H18734 to carry out analyses of parameters.

### Characterisation of raw edible oil wastewater

The physicochemical analyses of the raw, control and treated EORW samples were carried out using AOAC[40] methods.

### Photocatalytic degradation studies

Photocatalytic studies were performed with a slurry batch reactor. This reactor had a triple jacketed flow through a twin reactors system (Model: MS-H280-Pro). Lelesil Innovative Systems manufactured in India collaborate with the Small Scale Research Group, Faculty of Engineering, University of Benin, Benin City, Nigeria. The reactor system consists of two 5-L flow-through reactors. These are the primary reactor A, inside the photocatalytic reaction chamber, and the secondary reactor B outside. Also present are a peristaltic pump, hot plate with a magnetic stirrer, central jacket for UV lamps, and timer control digital clock. Hence, variables such as flow rate, temperature, agitation speed, UV irradiation, and irradiation time can be measured.[31]

The photocatalytic degradation studies were started by transferring the thoroughly mixed 0.25 g/L $TiO_2$-EORW solution into reactor B and connecting it to reactor A. The flow meter and magnetic stirrer were set at 100 mL/min and 900 rpm, respectively. We exposed reactor A to a 250-W mercury UV lamp, which was the source of UV light, and switched on the reactor system for 30 min. At the end of the reaction time, we collected the treated EORW samples, and centrifuged them at 5000 rpm for solid-liquid separation. A 200-mL supernatant was collected and used to carry out $BOD_5$ and COD analysis. A similar procedure was used for the control, which was EORW samples without the $TiO_2$ catalyst.

Moreover, we optimised the photocatalytic degradation process of EORW by determining the effects of the initial concentration of EORW, catalyst dose, agitation speed, and irradiation time on the photocatalytic degradation of EORW. The same procedure was followed in each case, varying one of the four variables each time: 100 mL/L or 250 mL/L, 0.25 g/L or 1.5 g/L, 300 rpm or 1500 rpm, and 20 min or 90 min, for initial concentration, catalyst dose, agitation speed, and irradiation time, respectively. The $BOD_5$ and COD for these were subsequently analysed at the end of each experiment.

### Analytical methods

We characterised the raw, control, and treated EORW samples by carrying out analysis of $BOD_5$, COD, total dissolved solids, dissolved oxygen, phenol, total suspended solids, $Cl^-$, $SO_4^{2-}$, oil and grease and $PO_4^{3-}$ using standard methods of water analysis.[41]

We calculated the reduction or performance efficiency ($E$) of pollutant removal from EORW using Equation 1:

$$E = \left[1 - \frac{cf}{ci}\right] X\ 100 \qquad \text{Equation 1}$$

where $C_f$ is final concentration, $C_i$ is initial concentration, and $E$ is reduction efficiency.

### Statistical analysis

Each experiment was done in triplicate and the mean and standard deviation (s.d.) of $n=3$ replicate results were recorded The data were analysed to determine significant differences using the Kruskal–Wallis H-test or one-way analysis of variance by ranks using Statistical Package for Social Sciences (SPSS) version 20 with a significance level of $p=0.05$.

## Results and discussion

### Characterisation of raw edible oil wastewater

The characteristics of the raw, control, and treated EORW are summarised in Table 1. The results show that the raw EORW was highly contaminated as expressed by its high COD, $BOD_5$, oil and grease, total suspended solids, $SO_4^{2-}$, etc. Table 1 shows that the physicochemical parameter values for the raw EORW were above or below the acceptable limits for discharged effluent specified by the Nigerian National Environmental Standard and Regulation Enforcement Agency (NESREA).[42] This could be because of the presence of chemical compounds such as phenols, heavy metals, catalysts, oxidisable substances, and fats and oils.[3,43] We observed a significant difference ($p<0.05$) between the raw or control physicochemical parameters and the treated EORW. However, the results showed no significant difference ($p<0.05$) in the physicochemical parameters of the raw and control EORW. Moreover, we observed that the photodegradation process reduced $BOD_5$, COD, total suspended solids, total dissolved solids, oil and grease, phenol, $SO_4^{2-}$, $PO_4^{3-}$, and $Cl^-$ in the treated EORW by 78.9%, 87.6%, 76.84%, 83.2%, 68.6%, 83.2%, 65.8% and 69.3%, respectively.


**Figure 1:** Set-up of the flow-through twin photocatalytic reactors system.

**Table 1:** Physicochemical characteristics of the raw and treated edible oil refinery wastewater

| Parameter | REORW | TEORW1 | TEORW2 | Limit | NESREA |
|---|---|---|---|---|---|
| pH | 2.8 ± 0.05 | 6.1 ± 1.3 (+54.1) | 3.6 ± 0.42 (+22.2) | 5.5–12 | 6–9 |
| Dissolved oxygen (mg/L) | 1.45 ± 0.03 | 3.7 ± 1.23 (+60.8) | 1.82 ± 0.3 (+20.3) | >3.0 | >5 |
| Biological oxygen demand ($BOD_5$) (mg/L) | 1620 ± 52.6 | 341.8 ± 18.3 (78.9) | 1252 ± 48.1 (22.7) | 50 | 50 |
| Chemical oxygen demand (COD) (mg/L) | 8265 ± 74.2 | 1074.5 ± 51.7 (87.0) | 7407.2 ± 92 (10.4) | 250 | 90 |
| BOD/COD | 0.20 | 0.32 | 0.17 | | |
| Oil/grease (mg/L) | 670.6 ± 27.6 | 218.62 ± 25.3 (67.4) | 454.3 ± 25 (32.3) | – | – |
| Total dissolved solids (mg/L) | 5920.7 ± 89.3 | 994.68 ± 63.2 (83.2) | 5364 ± 76.3 (9.4) | 400 | 10 |
| Total suspended solids (mg/L) | 1739.2 ± 71.6 | 402.8 ± 58.7 (76.84) | 1612 ± 63.5 (7.3) | 4000 | 500 |
| Phenols (mg/L) | 172.41 ± 4.8 | 54.14 ± 2.4 (68.6) | 139.7 ± 13.4 (19) | 1000 | 30 |
| $PO_4^{2-}$(mg/L) | 96.71 ± 8.4 | 33.08 ± 2.07 (65.8) | 68.05 ± 9.7 (29.6) | 150 | 0.5 |
| $SO_4^{2-}$(mg/L) | 2180.3 ± 75.12 | 366.24 ± 1.4 (83.2) | 1896.7 ± 76.3 (13) | 25 | |
| $Cl^-$(mg/L) | 574.63 ± 29.5 | 176.28 ± 8.4 (69.3) | 456.5 ± 36.7 (20.6) | 1500 | 250 |

*REORW, raw EROW; TEORW1, treated EORW using $TiO_2$; TEORW2, treated EORW without $TiO_2$ (control)*

*Results in mean±s.d.,* n=*3, and percentage reduction or improvement (+) in parentheses.*

The pH and dissolved oxygen increased by 54.1% and 60.8%, respectively. This suggests that the photodegradation process successfully removed suspended solids, extractable solvents, sulfates, and phosphates. The ratio of BOD/COD for raw and control EORW was 0.196 and 0.17, respectively. Aslan et al.[2] reported that vegetable oil wastewater with a low BOD/COD ratio (0.2) contained recalcitrant organic matter. However, Pintor et al.[44] and Welz et al.[45] reported a high BOD/COD ratio. They reported that vegetable oil wastewater contains high amounts (60–83%) of biodegradable organics; hence it does not require any pretreatment process. The ratio of BOD/COD for the treated EORW increased to 0.32. This shows that the photodegradation process increased its biodegradability; hence further biological treatment methods are highly encouraged. Liu et al.[18], Esteves et al.[15], Jamil et al.[46] and Xing et al.[47] reported similar results. They stated that photocatalytic pretreatment of excess sludge resulted in increased biodegradability of recalcitrant matter, allowing for further biological treatment of EORW. Also, we found that the physicochemical parameters of the treated EORW were within the acceptable water quality standards permissible for discharged effluent.[42]

## *Influence of operating parameters on the photodegradation of edible oil wastewater*

### Effect of initial concentration

Figure 2 shows the effect of the initial concentration of EORW on reduction efficiencies of $BOD_5$ and COD. COD and $BOD_5$ increased from 61.8% to 76.4% and 61.8% to 76.4%, respectively, as the initial

concentration of EORW reduced from 250 mL/L to 100 mL/L. Therefore, the optimum initial concentration was 100 mg/L. The lower efficiencies witnessed at higher initial concentrations of EORW could be because of the formation of several layers of adsorbed pollutant molecules on the photocatalyst ($TiO_2$) surface which inhibits the organic molecules' reaction with the holes or hydroxyl and oxygen free radicals (•OH and $O^{2-}$) as the molecules have limited access to these radicals.[48] Aisien et al.[31] reported similar results in applying periwinkle shell ash as a photocatalyst for the degradation of naphthalene in an aqueous solution. Therefore, if the initial concentration is increased, the relative ratio of hydroxyl radicals attacking the pollutant decreases, thereby decreasing photocatalytic efficiency.[49]

### Effect of irradiation time

Figure 3 shows the effect of irradiation time on the reduction efficiency of $BOD_5$ and COD during photodegradation of EORW. We observed a rapid photodegradation of EORW in the first 40 min of the process, as shown by the steep increase in the reduction efficiency of $BOD_5$ and COD. The photodegradation rate continued to increase for another 30 min but less rapidly, and the profile levels off after that, showing that an equilibrium had been reached. At equilibrium, the active sites on the $TiO_2$ particles are occupied by the pollutant molecules, which leads to saturation, hence there is no appreciable increase in photodegradation of EORW. The rapid rate of photodegradation observed in the initial stage of the process may be attributed to the abundant availability of active sites on the surface of the $TiO_2$ photocatalyst. These sites are later occupied by the pollutant molecules as the process progresses, which results in a slowed photodegradation rate in the latter part of the process. Other researchers[18,31,48] have reported similar results for other wastewater.

### Effect of catalyst dose

The effect of catalyst dose on the reduction efficiency of $BOD_5$ and COD during photodegradation of EORW in the presence of the $TiO_2$ catalyst is shown in Figure 4. We observed that, using $TiO_2$ as a photocatalyst, the reduction efficiencies of $BOD_5$ and COD initially increased with an increase in catalyst dose up to maximum values of 70.8% and 78% for $BOD_5$ and COD, respectively, at a catalyst dose of 1.25 g/L. Further increases in catalyst dose resulted in an insignificant increase in reduction efficiencies of $BOD_5$ and COD. This might be attributed to the initial increase in reduction efficiency to lower doses of the catalyst. The catalyst surface and its light absorption are the limiting factors. Hence increasing the catalyst loading enhances the efficiency of the process. This is because of the increase in the number of active sites on the photocatalyst surface, which increases the number of free radicals (•OH and $O^{2-}$) produced in EORW.[50] Aisien et al.[31] reported similar observations in the photodegradation of aniline using periwinkle shell ash as a photocatalyst. Suri et al.[51] reported that using a higher dosage of the catalyst might not be advisable.

### Effect of agitation speed

The effect of the agitation speed on the reduction efficiencies of $BOD_5$ and COD during the photodegradation EORW is shown in Figure 5. We observed that $BOD_5$ and COD increased from 65% to 72.2% and 70% to 75%, respectively, as the agitation speed increased from 300 rpm to 900 rpm. Further increases in agitation speed resulted in decreases in $BOD_5$ and COD. Compared with the effect of irradiation time and catalyst dose, agitation speed had the lowest influence. The optimum agitation speed required for photocatalytic degradation of EORW is 900 rpm. This finding is supported by Wu et al.[52] They stated that the degradation rate increases slightly with agitation speed. The effect of agitation speed can be explained in two ways. Firstly, it is well known that two steps in series govern heterogeneous photocatalysis. Then the overall mass transfer is the summation of the mass-transfer resistance and the chemical reaction resistance. So the increase of the agitation speed leads to a high mass transfer and high degradation rate of pollutants. Secondly, the agitation speed increase can promote oxygen transfer in the liquid phase and thereby increase the degradation kinetics.[53]

## *Kinetics of photocatalytic degradation of edible oil wastewater*

We studied the photodegradation process kinetics using pseudo-first-order, pseudo-second-order, and intra-particle diffusion kinetic models.

### Pseudo-first-order model

The pseudo first-order equation is expressed as follows[31]:

$$r = -\frac{dC}{dt} kC \qquad \text{Equation 2}$$

The pseudo-first-order reaction rate constant, $k$, was obtained from the integrated linear form of Equation 2 as follows:

$$k = \frac{\ln C_o/C_i}{t} \qquad \text{Equation 3}$$

where $C_o$ is the initial concentration of EORW, and $k$ ($min^{-1}$) is the pseudo-first-order rate constant.

A plot of ln $C_o/C_i$ versus time ($t$) resulted in a linear relationship from which $k_1$ and $R^2$ were determined from the slope of the graph as shown in Figure 6. The values of the constants are given in Table 2. We observed a linear relationship in Figure 6, which suggests that the pseudo-first-order equation applies to the photodegradation process. Moreover, we observed that the values of $k$ decreased with an increase in initial EORW concentration. This agrees with the inference made from the results presented in Figure 2 that increasing the initial EORW concentration leads to a reduction in photodegradation reaction rate. According to Daneshvar et al.[54], a reduction in inactive sites on the catalyst surface may be the reason for the decrease in the rate constant.

### Pseudo-second-order model

The pseudo-second-order kinetic model is expressed in its integrated linear form as follows[55]:

$$\frac{t}{q_t} = \frac{1}{k_2 q_e^2} + \frac{1}{q_e} t \qquad \text{Equation 4}$$

$$h = k_2 q_e^2 \qquad \text{Equation 5}$$

where $k_2$ is the rate constant of the pseudo-second-order process (g/mg.min).

The plot of ($t/qt$) versus $t$ as shown in Figure 7 was used to calculate the values of the model parameters. The kinetic constants calculated are shown in Table 2. High $R^2$ values for the pseudo-second-order model can be seen from Figure 7. The model could describe the photodegradation process kinetics for initial EORW concentration values of 100 mg/L and 150 mg/L. Beyond these initial concentrations, the model failed to describe the mechanism of the process. However, it is pronounced from Figure 7 that the pseudo-second-order model was the best to represent the kinetics of the photodegradation process, as stated in approximate 1 $R^2$ values (Table 2).

Comparing the pseudo-first-order and the pseudo-second-order kinetic models. We observed that both could successfully describe the kinetic of the photodegradation process of EORW. However, the $R^2$ values for the pseudo-second-order are nearer to 1. It means that the pseudo-second-order model gave the best fit. This suggests that the pseudo-second-order model better represented the photodegradation of EORW than did the pseudo-first-order model.

### Intra-particle diffusion model

The process's diffusion mechanism was modelled using the intra-particle diffusion model as follows[56]:

$$q_t = K_p t^{1/2} + C \qquad \text{Equation 6}$$

where $K_p$ is the intra-particle diffusion rate constant (mg/g.$min^{1/2}$) and $C$ is a measure of the boundary layer effect.

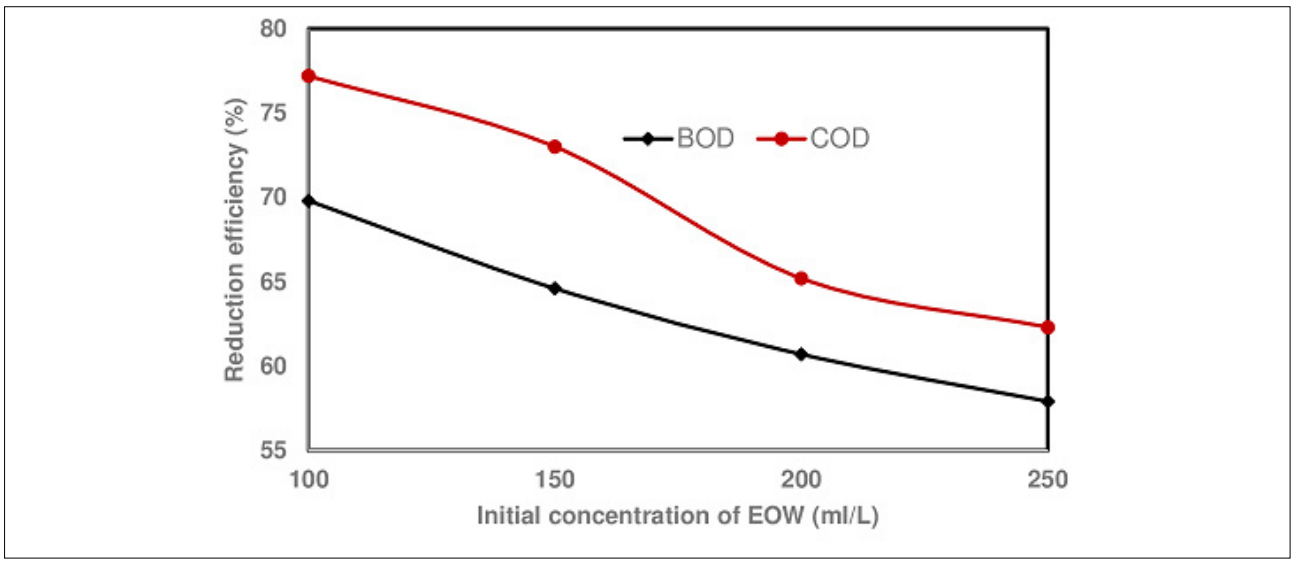



**Figure 2:** Effect of the initial concentration of the edible oil wastewater (EOW) on the reduction efficiency of biological oxygen demand ($BOD_5$) and chemical oxygen demand (COD).

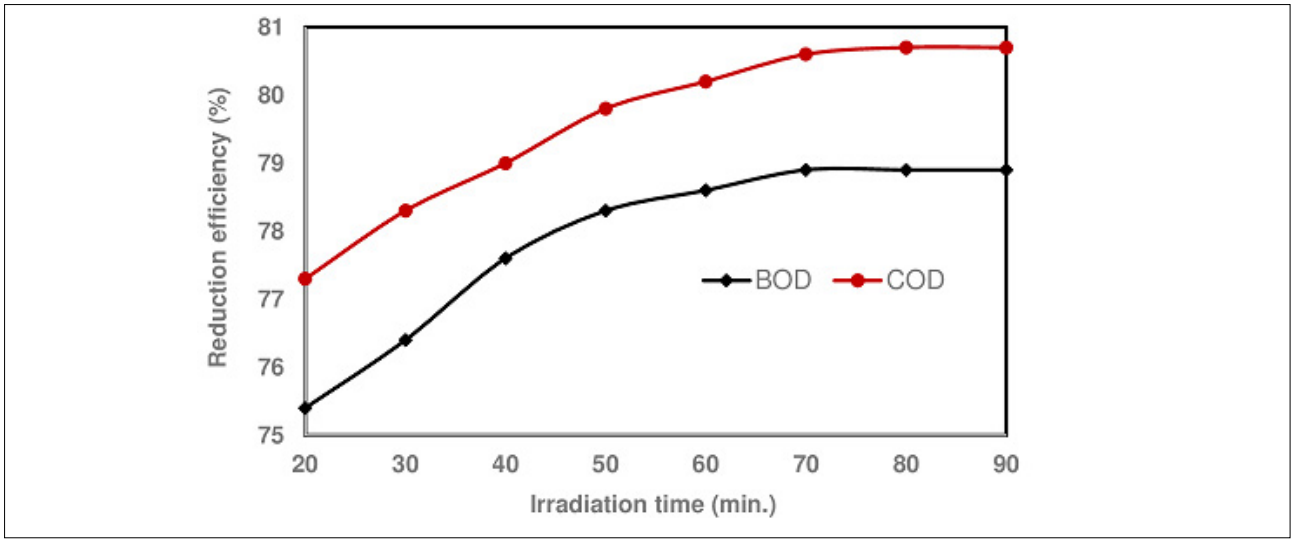



**Figure 3:** Effect of irradiation time on the reduction efficiency of biological oxygen demand ($BOD_5$) and chemical oxygen demand (COD).

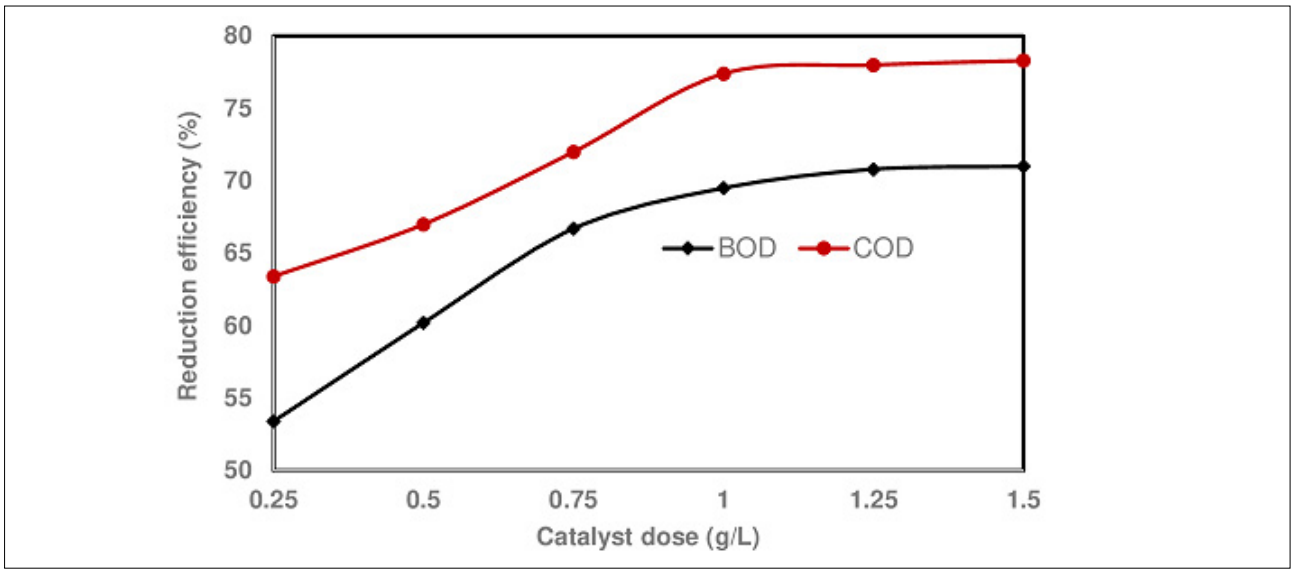



**Figure 4:** Effect of catalyst dose on the reduction efficiency of biological oxygen demand ($BOD_5$) and chemical oxygen demand (COD).

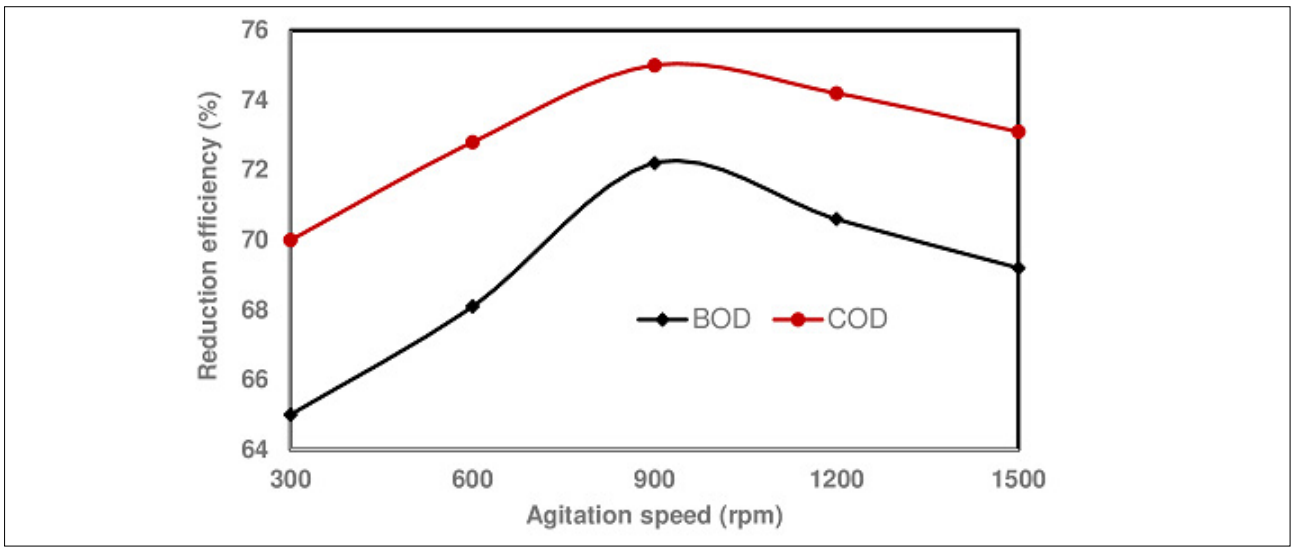



**Figure 5:** Effect of agitation speed on the reduction efficiency of biological oxygen demand ($BOD_5$) and chemical oxygen demand (COD).

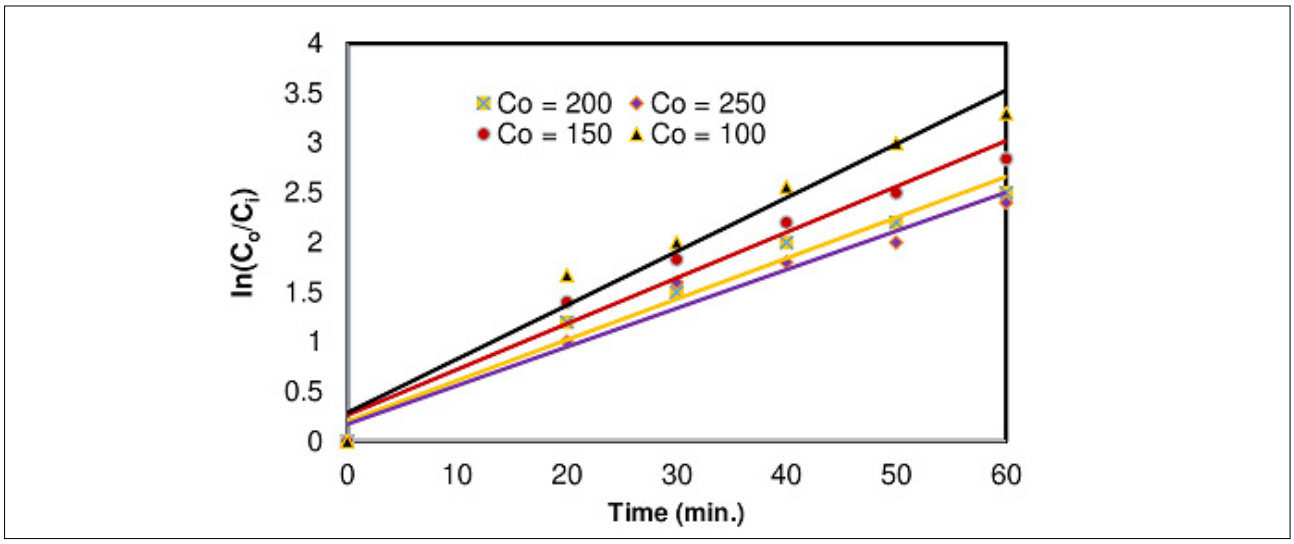



**Figure 6:** Pseudo-first-order model fitted to batch equilibrium data for edible oil wastewater photodegradation.

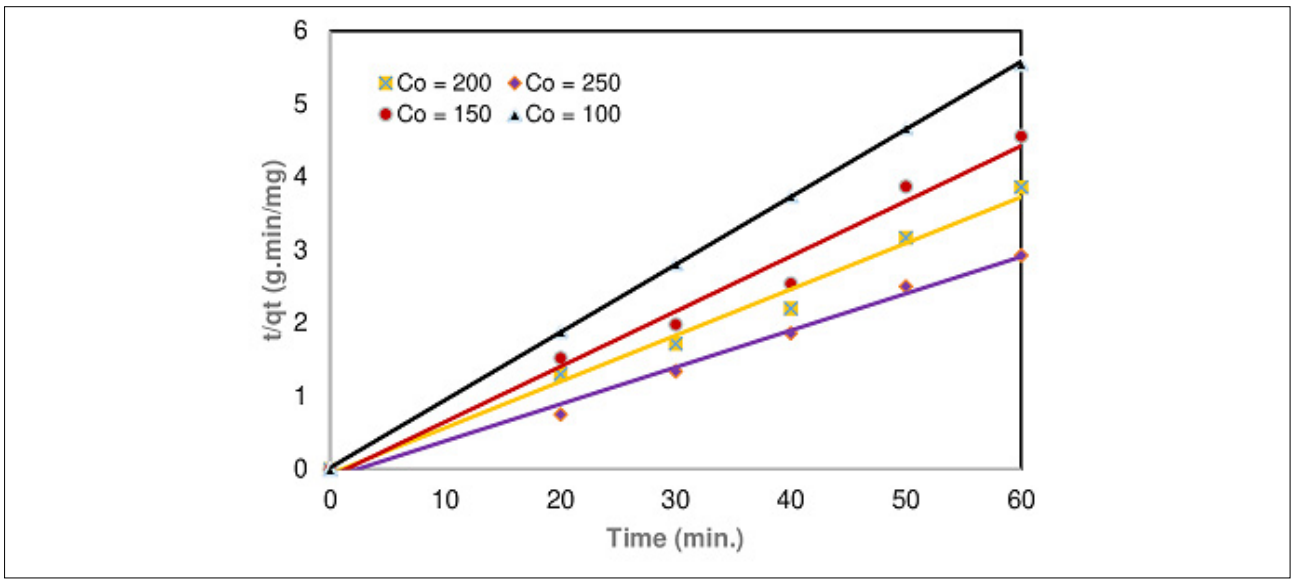



**Figure 7:** Pseudo-second-order model fitted to batch equilibrium data for edible oil wastewater photodegradation.

The value of $C$ shows the contribution of the surface sorption to the rate-controlling step. According to the model, a plot of $q_t$ versus $t^{1/2}$ should be a straight line from the origin if the adsorption mechanism follows the intra-particle diffusion only. The values of the intra-particle diffusion model parameters are presented in Table 2. The plot of $q_t$ versus $t^{1/2}$ from which we used to calculate the parameters is shown in Figure 8. We observed from Figure 8 that the straight-line plots for the different initial concentrations of EORW investigated did not begin from the origin. In addition, the $R^2$ values were lower than shown in Table 2 and Figure 8. The plot shows the existence (although not significant) of some boundary layer effect and further showed that intra-particle diffusion was not the only rate-limiting step. The lower $R^2$ values obtained for the kinetic parameters within the concentration range investigated, show that the intra-particle diffusion model was not able to fully describe the diffusion mechanism of the process.

## Conclusions

The following conclusions can be drawn from this study:

1. The characterisation of raw EORW shows it was highly polluted because the physicochemical parameters were outside the acceptable water quality standards for discharged effluent specified by NESREA as maximum permissible limits.
2. The photocatalytic treatment of EORW gave reduction efficiencies ranging between 65.8% ($PO_4^{3-}$) and 87.0% (COD) and the improvement in efficiency was 54.1% (pH) and 60.8% (dissolved oxygen).
3. Photocatalytic treatment of EORW gave a negligible reduction efficiency in the absence of $TiO_2$ as a photocatalyst.
4. The biodegradability of EORW increased from 0.196 to 0.32 after photocatalytic treatment.
5. Optimum values for the process variables were an EORW initial concentration of 100 mL/L, catalyst dosage of 1.25 g/L, agitation speed of 900 rpm, and irradiation time of 50 min.
6. The catalyst dosage, agitation speed, and irradiation effects on the photocatalytic treatment of EORW were in the order irradiation time > catalyst dosage > agitation speed.
7. The pseudo-first-order and pseudo-second-order kinetic models could describe the kinetics of pollutant removal from EORW using photocatalytic degradation. However, the pseudo-second-order kinetic model gave the best fit, with the highest $R^2$ value ($R^2 > 0.98$).
8. The diffusion mechanism is not fully described by the intra-particle diffusion model.
9. The photocatalytic treatment is more effective, cheaper and saves time in comparison to other methods such as chemical and biological methods because of its quick and efficient removal and elimination of organic matter.

## Acknowledgements

We gratefully acknowledge the late Isaac Ayoku Uwagboe for his supervision of the design and fabrication of the photocatalytic reactor system and for his kind provision of the reactor system.

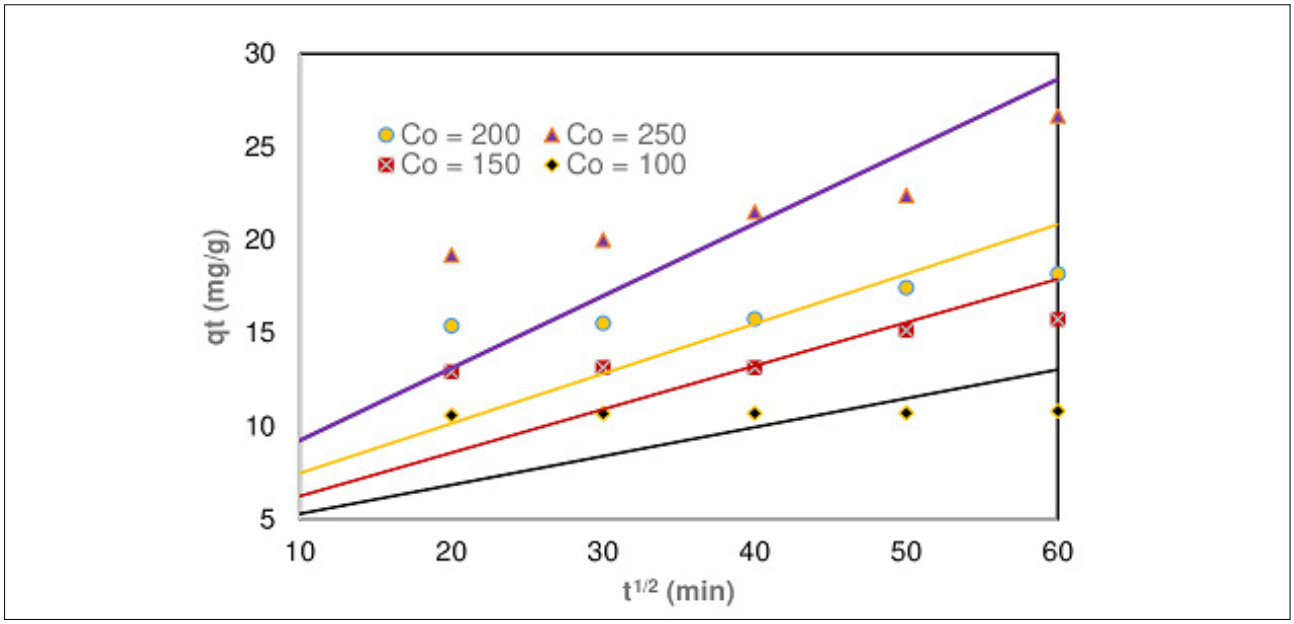



**Figure 8:** Intra-particle diffusion model fitted to batch equilibrium data for edible oil wastewater photodegradation.

**Table 2:** Constant parameter values for the photocatalytic degradation kinetics of edible oil refinery wastewater (EORW)

| Initial concentration of EORW (mg/L) | Pseudo-first-order model | | Pseudo-second-order model | | | Intra-particle-diffusion model | | |
|---|---|---|---|---|---|---|---|---|
| | $k_1$ | $R^2$ | $k_2$ | $q_e$ | $R^2$ | $K_p$ | $C$ | $R^2$ |
| 100 | 0.039 | 0.965 | 0.034 | 4.124 | 1.000 | 0.082 | 0.03 | 0.806 |
| 150 | 0.023 | 0.961 | 0.009 | 6.733 | 0.981 | 0.357 | 0.178 | 0.718 |
| 200 | 0.027 | 0.968 | 0.0045 | 9.560 | 0.987 | 0.675 | 0.342 | 0.742 |
| 250 | 0.018 | 0.964 | 0.052 | 16.785 | 0.992 | 0.902 | 0.456 | 0.587 |

## Competing interests

We have no competing interests to declare.

## Authors' contributions

E.T.A. contributed to the development or design of the methodology and creation of models and conducted the research and investigation process. She was involved in the preparation and creation of the published work, specifically writing, critically reviewing and revising the paper. F.A.A. was responsible for student supervision and oversight of the study. He contributed to the research activity planning and execution, ensured the preparation and creation of the published work, and critically reviewed and revised the paper.

## References

1. Davarnejad R, Sabzehei M, Parvizi F, Heidari S, Rashidi A. Study on soybean oil plant wastewater treatment using the electro-Fenton technique. Chem Eng Technol. 2019;42(12):56–63. https://doi.org/10.1002/ceat.201800765
2. Aslan S, Alyuz B, Bozkurt Z, Bakaoglu M. Characterization and biological treatability of edible oil wastewaters. Pol J Environ Stud. 2009;18:533–538.
3. Velioglu SG, Curi KY, Camililar SR. Activated sludge treatability of olive oil-bearing wastewater. Water Res. 1992;25(10):1415–1420. https://doi.org/10.1016/0043-1354(92)90135-q
4. Boyer MJ. Current pollution control practices in the United States. J Am Oil Chem Soc. 1984;61(5):297–309.
5. Tehrani GM, Mahnoudi MM, Borgheipour H, Nezampour A. Evaluation of the efficiency of integrated fixed-film activated sludge reactor for treatment of wastewater from vegetable oil industries. Arch Hyg Sci. 2018;7(3):192–199.
6. Niazmand R, Jahani M, Kalantarian S. Treatment of olive processing wastewater by electrocoagulation: An effectiveness and economic assessment. J Environ Manage. 2019;248, Art. #109262. https://doi.org/10.1016/j.jenvman.2019.109262
7. Sharma S, Simsek H. Treatment of canola-oil refinery effluent using electrochemical methods: A comparison between combined electrocoagulation þ electrooxidation and electrochemical peroxidation methods. Chemosphere. 2019;221:630–639.https://doi.org/10.1016/j.chemosphere.2019.01.066
8. Sharma S, Sharma AK, Verma S, Dodiya HS. Treatment of edible oil refinery wastewater by using a chemical and biological process. Int J Eng Sci Res Technol. 2014;3:115–119.
9. Un UT, Koparal AS, Ogutveren UB. Electrocoagulation of vegetable oil refinery wastewater using aluminum electrodes. J Environ Manage. 2009;90:428–433. https://doi.org/10.1016/j.jenvman.2007.11.007
10. Dkhissi O, El Hakmaoui A, Souabi S, Chatoui M, Jada A, Akssira M. Treatment of vegetable oil refinery wastewater by coaulation-flocculation using the cactus as a bio-flocculant. Mater Environ Sci. 2018;9:18–25. https://doi.org/10.26872/jmes.2018.9.1.3
11. Šereš Z, Maravič N, Takači A, Nicolič I, Soronja-Simovič D, Jokič A, et al. Treatment of vegetable oil refinery wastewater using alumina ceramic membrane: Optimization using response surface methodology. J Cleaner Prod. 2016;112:3132–3137. https://doi.org/10.1016/j.jclepro.2015.10.070
12. Pazdzior K, Bilinska L, Ledakowicz S. A review of the existing and emerging technologies in the combination of AOPs and biological processes in industrial textile wastewater treatment. Chem Eng J. 2019;376, Art. #120597. https://doi.org/10.1016/j.cej.2018.12.057
13. Ma A, Lei T, Ji X, Gao X, Gao C. Submerged membrane bioreactor for vegetable oil wastewater treatment. Chem Eng Technol. 2015;38:101–109. https://doi.org/10.1002/ceat.201400184
14. Kalat DG, Yüceer A. Anaerobic mesophilic and thermophilic treatability of vegetable oil refining wastewater. Process Saf Environ Prot. 2017;109:151–157. https://doi.org/10.1016/j.psep.2017.04.001
15. Esteves BM, Rodrigues CSD, Madeira LM. Synthetic olive mill wastewater treatment by Fenton's process in batch and continuous reactors operation. Environ Sci Pollut Res. 2018;25:34826–34838. https://doi.org/10.1007/s11356-017-0532-y
16. Chen X, Mao S. Titanium dioxide nanomaterials: Synthesis, properties, modifications, and applications. Chem Rev. 2007;107:2891–2959. https://doi.org/10.1021/cr0500535
17. Setvin M, Shi X, Hulva J, Simschitz T, Parkinson GS, Schmid M. Methanol on anatase $TiO_2$ (101) mechanistic insights into photocatalysis. ACS Catal. 2017;7(10):7081–7091. https://doi.org/10.1021/acscatal.7b02003
18. Liu C, Shi W, Kim M, Yang Y, Lei Z, Zhang Z. Photocatalytic pretreatment for the redox conversion of waste activated sludge to enhance biohydrogen production. Int J Hydrogen Energy. 2013;38:7246–7252. https://doi.org/10.1016/j.ijhydene.2013.03.147
19. Welz PJ. Edible seed oil waste: Status quo and future perspectives. Water Sci Technol. 2019;80(11):2107–2116. https://doi.org/10.2166/wst.2020.043
20. Hu X, Hu X, Tang C, Wen S, Wu X, Long J, et al. Mechanisms underlying degradation pathways of microcystin-LR with doped photocatalysis. Chem Eng J. 2017;330:335–371.
21. Poza-Nogueiras V, Rosales E, Pazos M, Sanroma MA. Current advances and trends in electron-Fenton process using heterogeneous catalysts: A review. Chemosphere. 2018;201:399–416. https://doi.org/10.1016/j.chemosphere.2018.03.002
22. Ahmed SN, Haider W. Heterogeneous photocatalysis and its potential applications in water and wastewater treatment: A review. Nanotechnology. 2018;29(34), Art. #342001. https://doi.org/10.1088/1361-6528/aac6ea
23. Lee SY, Park SJ. $TiO_2$ photocatalyst for water treatment applications. J Ind Eng Chem. 2013;19:1761–1769.
24. M'Arimi MM, Mecha CA, Kiprop AK, Ramkat R. Recent trends in applications of advanced oxidation processes (AOPs) in bioenergy production: Review. Renew Sust Energ Rev. 2020; 121, 109669. https://doi.org/10.1016/j.rser.2019.109669
25. Chen B, Nam SN, Westerhoff PK, Krasner SW, Amy G. Fate of effluent organic matter and DBP precursors in an effluent-dominated river: A case study of wastewater impact on downstream water quality. Water Res. 2009;43:1755–1765. https://doi.org/10.1016/j.watres.2009.01.020
26. Hur J, Lee TH, Lee BM. Estimating the removal efficiency of refractory dissolved organic matter in wastewater treatment plants using a fluorescence technique. Environ Technol. 2011;32:1843–1850. https://doi.org/10.1080/09593330.2011.565078
27. Dhanke P, Wagh S. Treatment of vegetable oil refinery wastewater with biodegradability index improvement Mater Today Proc. 2020;27(1):181–187. https://doi.org/10.1016/j.matpr.2019.10.004
28. Nagar N, Devra V. A kinetic study on the degradation and biodegradability of silver nanoparticles catalyzed methyl orange and textile effluents. Heliyon 2019;5(3), e01356. https://doi.org/10.1016/j.heliyon.2019.e01356
29. Gilja V, Vrban I, Mandic V, Žic M, Murgic ZH. Preparation of a PANI/ZnO composite for efficient photocatalytic degradation of acid blue. Polymer. 2018;10:940–957. https://doi.org/10.3390/polym10090940
30. Zhou Z, Liu X, Sun K, Lin C, Ma J, He M, et al. Persulfate-based advanced oxidation processes (AOPs) for organic-contaminated soil remediation: A review. Chem Eng J. 2019;372:836–851. https://doi.org/10.1016/j.cej.2019.04.213
31. Aisien F, Amenaghawon A, Mededode A. Heterogeneous photocatalytic degradation of naphthalene using periwinkle shell ash: Effect of operating variables, kinetic and isotherm study. S Afr J Chem Eng. 2014;9(1):31–45.
32. Omri A, Benzina M. Almond shell activated carbon: Adsorbent and catalytic support in the phenol degradation. Environ Monit Assess. 2014;186(6):3875–3890. https://doi.org/10.1007/s10661-014-3664-2
33. Hao D, Yang Z, Jiang C, Zhang J. Photocatalytic activities of $TiO_2$ coated on different semiconductive SiC foam supports. J Mater Sci Technol. 2013;29(11):1074–1078. https://doi.org/10.1016/j.jmst.2013.08.021
34. Kalebaila KK, Fairbridge C. UV photocatalytic degradation of commercial naphthenic acid using $TiO_2$-zeolite composites. J Water Res Prot. 2014;6(12):1198–1206. https://doi.org/10.4236/jwarp.2014.612109
35. Bel Hadjltaief H, Galvez ME, Da Costa P, Ben Zina M. $TiO_2$/clay as a heterogeneous catalyst in photocatalytic/photochemical oxidation of anionic reactive blue 19. Arabian J Chem. 2019;12(7):1454–1462. https://doi.org/10.1016/j.arabjc.2014.11.006

36. Ao Y, Xu J, Fu D, Shen X, Yuan C. Low temperature preparation of anatase $TiO_2$-coated activated carbon. Colloids Surf A Physicochem Eng Asp. 2008;312(2–3):125–130. https://doi.org/10.1016/j.colsurfa.2007.06.039

37. Omri A, Benzina M, Bennour F. Industrial application of photocatalysts prepared by hydrothermal and sol-gel methods. J Ind Eng Chem. 2015;21:356–362. https://doi.org/10.1016/j.jiec.2014.02.045

38. Takeuchi M, Kimura T, Hidaka M, Rakhmawaty D, Anpo M. Photocatalytic oxidation of acetaldehyde with oxygen on $TiO_2$/ZSM-5 photocatalysts: Effect of hydrophobicity of zeolites. J Catal. 2007;246(2):235–240. https://doi.org/10.1016/j.jcat.2006.12.010

39. Takeuchi M, Hidaka M, Anpo M. Efficient removal of toluene and benzene in gas phase by the $TiO_2$/Y-zeolite hybrid photocatalyst. J Hazard Mater. 2012;237:133–139. https://doi.org/10.1016/j.jhazmat.2012.08.011

40. AOAC. Official methods of analysis of the association of official analytical chemists. 19th ed. Rockville, MD: AOAC; 2012. p. 234–256.

41. American Public Health Association (APHA). Standard methods for the examination of water and wastewater (22nd). Washington DC: APHA; 2012.

42. Nigerian National Environmental Standard and Regulation Enforcement Agency (NESREA). Guidelines and standards for water quality in Nigeria. Abuja: NESREA; 2008.

43. Chipasa KB. Limits of physicochemical treatment of wastewater in the vegetable oil refining industry. Pol J Environ Stud. 2001;10:141–147.

44. Pintor AMA, Vilar VJP, Botleho CMS, Boaventura RAR. Optimization of a primary gravity separation treatment for vegetable oil refinery wastewater. Clean Technol Environ Policy. 2014;16:1725–1734. https://doi.org/10.1007/s10098-014-0754-3

45. Welz PJ, Le Roes-Hill M, Swartz CD. NATSURV 6: Water and wastewater management in the edible oil industry. 2nd ed. WRC TT 702/16. Pretoria: Water Research Commission; 2017.

46. Jamil TS, Ghaly MY, El-Seesy IE, Souaya ER, Nasr RA. A comparative study among different photochemical oxidation processes to enhance the biodegradability of paper mill wastewater. J Hazard Mater. 2011;185:353–358. https://doi.org/10.1016/j.jhazmat.2010.09.041

47. Xing R, Zheng Z, Wen D. Comparison between UV and VUV photolysis for the pre- and post-treatment of coking wastewater. J Environ Sci. 2015;29:45–50. https://doi.org/10.1016/j.jes.2014.10.003

48. Abdollahi Y, Abdullah AH, Zainal Z, Yusof NA. Photodegradation of m-cresol by zinc oxide under visible-light irradiation. Int J Chem. 2011;3(3):31–43. https://doi.org/10.5539/ijc.v3n3p31

49. Kumar DP, Mohamed AR, Bhatia S. Wastewater treatment using photocatalysis: Destruction of methylene blue dye from wastewater streams. J Kejuruteraan. 2002;14:17–30.

50. Konstantinou IK, Albanis TA. $TiO_2$-assisted photocatalytic degradation of azo dyes in aqueous: Kinetic and mechanistic investigation: A review. Appl Catal B Environ. 2004;49:1–14. https://doi.org/10.1016/j.apcatb.2003.11.010

51. Suri R, Liu J, Hand D, Crittenden J, Perram D, Mullims M. Heterogeneous photocatalytic oxidation of hazardous organic contaminants in water. Water Env Res. 1993;65(5):665–673. https://doi.org/10.2175/wer.65.5.9

52. Wu CH, Chang HW, Chem JM. Basic dye decomposition kinetics in a photocatalytic slurry reactor. J Hazard Mater B. 2006;137:336–343. https://doi.org/10.1016/j.jhazmat.2006.02.002

53. Merabet S, Bouzazab A, Wolbert D. Photocatalytic degradation of indole in a circulating upflow reactor by UV/$TiO_2$ process-Influence of some operating parameters. J Hazard Mater. 2009;166:1244–1249. https://doi.org/10.1016/j.jhazmat.2008.12.047

54. Daneshvar N, Aber S, Dorraji MS, Khataee AR, Rasoulifard MH. Preparation and investigation of the photocatalytic properties of ZnO nanocrystals: Effect of operational parameters and kinetic study. Int J Chem Nucl Metall Mater Eng. 2008;1(5):66–71.

55. Ho YS. Review of second-order models for adsorption systems. J Hazard Mater. 2006;B136:681–689. https://doi.org/10.1016/j.jhazmat.2005.12.043

56. Weber WT, Morris JC. Kinetics of adsorption on carbon from solution. J Sanitary Eng. 1963;89(2):31–60.

**AUTHORS:**
Louisa C. Sarkodie[1]
Philomena Entsie[2]
Mariam E. Boakye-Gyasi[1]
Frederick W.A. Owusu[1]
Marcel T. Bayor[1]
Kwabena Ofori-Kwakye[1]

**AFFILIATIONS:**
[1]Department of Pharmaceutics, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana
[2]Department of Herbal Medicine, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana

**CORRESPONDENCE TO:**
Mariam Boakye-Gyasi

**EMAIL:**
melboakye-gyasi.pharm@knust.edu.gh

**DATES:**
**Received:** 22 June 2021
**Revised:** 16 Sep. 2021
**Accepted:** 20 Sep. 2021
**Published:** 29 Nov. 2021

**HOW TO CITE:**
Sarkodie LC, Entsie P, Boakye-Gyasi ME, Owusu FWA, Bayor MT, Ofori-Kwakye K. Evaluation of the binding and disintegrating properties of gum obtained from the stem bark of *Cinnamomum zeylanicum*. S Afr J Sci. 2021;117(11/12), Art. #11550. https://doi.org/10.17159/sajs.2021/11550

**ARTICLE INCLUDES:**
☒ Peer review
☐ Supplementary material

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

**EDITOR:**
Teresa Coutinho

**KEYWORDS:**
natural excipients, *Cinnamomum zeylanicum* gum, binder, disintegrant, paracetamol tablets

**FUNDING:**
None

# Evaluation of the binding and disintegrating properties of gum obtained from the stem bark of *Cinnamomum zeylanicum*

Excipients are the various ingredients, apart from the active pharmaceutical ingredients, which are added to pharmaceutical formulations. Excipients obtained from natural sources are preferred over those from synthetic sources because they are cheap, biocompatible and readily available. Gums are made up of carbohydrate units which are linked by glycosidic bonds. This study was aimed at evaluating the potential binding and disintegrating properties of gum obtained from the bark of *Cinnamomum zeylanicum*, which was obtained from Effiduase in the Ashanti region of Ghana. The gum was extracted using 96% ethanol and the moisture content, Fourier transform infrared spectroscopy spectra, water holding capacity, swelling index and flow properties of the gum were determined. The gum was used to formulate tablets at different concentrations (10% $^{w}/_{v}$, 15% $^{w}/_{v}$ and 20% $^{w}/_{v}$) as binder with acacia as the standard. The gum was also used to formulate tablets at different concentrations (5% $^{w}/_{v}$, 7.5% $^{w}/_{v}$ and 10% $^{w}/_{v}$) as disintegrant with starch as the standard. Quality control tests were then conducted on all formulated tablets. The gum exhibited good flow and physicochemical properties. All formulated tablets passed the uniformity of weight test, friability test, disintegration test, hardness test, uniformity of dimensions test and drug content. All batches of tablets, except Batch 7, passed the dissolution test. Based on the study carried out, *C. zeylanicum* gum can be used as an alternative excipient to acacia and starch as a binder and a disintegrant, respectively.

**Significance:**

- A natural polysaccharide (gum) from the bark of *Cinnamomum zeylanicum* tree can be harnessed and commodified as a pharmaceutical excipient (binder and disintegrant) in the production of immediate release tablets.

## Introduction

Pharmaceutical excipients constitute about 90% of dosage forms and can be defined as any substance added to the active pharmaceutical ingredient(s) during the process of pharmaceutical manufacturing.[1] In pharmaceutical formulations, excipients may be employed as binders and disintegrants in immediate release tablets. In formulating immediate release tablets, binders promote the cohesiveness of powder mixtures, improving the flow of granules and the strength of the resulting tablet.[2] Disintegrants are included in the formulation of tablets to facilitate their break-up into smaller fragments in an aqueous medium. Binders and disintegrants may be from natural, synthetic or semi-synthetic sources. The quest for natural excipients has increased tremendously over the past decade. Pharmaceutical excipients obtained from natural origin are known to be economical, biodegradable, safe and relatively abundant.[3] In pharmaceutical industries, natural polysaccharides like mucilages, starches and gums are used as fillers, binders and disintegrants.[4]

Gums are pathological products of plants resulting from the breakdown of cell walls as a result of plant injury or unfavourable environmental conditions. They are hydrophilic, amorphous, colloidal and sticky in nature with uronic and sugar units. They may be natural, synthetic or semi-synthetic in nature. Gums from okra, khaya, acacia and gellen, which are naturally occurring, have been commonly used as binders in tablet formulation.[5] The bark of *Cinnamomum zeylanicum* (Family: Lauraceae), aside from its medicinal properties as an anthelmintic and anti-inflammatory agent, has not been widely researched as a potential excipient in solid dosage forms.[6,7] *C. zeylanicum,* or Ceylon cinnamon tree, is readily available in the tropics and subtropics with Ghana being no exception. Gum obtained from the bark of the plant can be locally harnessed and investigated as a potential excipient in tablets. In this study, gum obtained from the bark of *C. zeylanicum* was investigated for its potential as both a binder and disintegrant in the formulation of oral immediate release tablets.

## Materials and methods

### *Materials*

The bark of *C. zeylanicum* was obtained from Effiduase in the Ashanti region, Ghana, and authenticated at the Department of Herbal Medicine, Faculty of Pharmacy and Pharmaceutical Sciences, Kwame Nkrumah University of Science and Technology (Kumasi, Ghana) by a botanist. It was given a specimen voucher number of KNUST/HM1/2018/SB002. Diethyl ether, ethanol (96%), paracetamol powder, acacia powder, starch powder, talc, magnesium stearate and lactose were obtained from UK Chemicals, Kumasi. All other reagents used were of analytical grade.

### *Extraction of gum from* Cinnamomum zeylanicum *bark*

Distilled water (7 L) was added to 200 g *C. zeylanicum* powdered bark and allowed to stand for 24 h at room temperature (27 °C). The mixture was then boiled for 15 min, allowed to cool and filtered with a calico strainer to remove any debris to obtain the gum. Ethanol (96%) was used in purifying the gum obtained via precipitation. The precipitated gum was filtered, washed with diethyl ether and dried in a hot air oven at 60 °C for 12 h.[8] The dried gum was milled, passed through a sieve with an aperture size 180 $\mu$m and appropriately stored for use.

### *Determination of moisture content*

A mass of 1 g of the gum was transferred into a dried crucible of known weight and placed in an oven at 105 °C for 5 h. The crucible containing the gum was removed and allowed to cool at room temperature and weighed.[9]

### *Swelling index and water holding capacity of* Cinnamomum zeylanicum *gum*

An amount of 0.5 g *C. zeylanicum* gum (CZG) was transferred into a measuring cylinder, tapped and its volume recorded as $V_0$. It was then dispersed in 10 mL of distilled water and allowed to stand for 24 h. The volume occupied by the gum was observed after 24 h ($V_{24}$) and calculated as:

$$\text{Swelling index} = \frac{V_{24} - V_0}{V_0} \times 100$$

Contents in the measuring cylinder used for determination of the swelling index were also used to determine the water holding capacity as described by Ofori-Kwakye's group.[10]

### *Determination of the pH of CZG*

*Cinnamomum zeylanicum* gum (0.1 g) was dispersed in 10 mL of water and stirred for complete dissolution. The pH was then checked with a pH meter (MW101, Milwaukee Instruments, Rocky Mount, NC, USA).[11] This was done in triplicate.

### *Compatibility studies on formulated granules*

Paracetamol powder, CZG and formulated granules were individually scanned using a Fourier transform infrared spectrophotometer (Bruker Alpha II, Germany) over 500–3500 $cm^{-1}$ wavelengths. Their spectra were then superimposed to assess the presence or absence of principal bands of paracetamol in the formulated tablets.

### *Determination of the flow property of CZG*

*Cinnamomum zeylanicum* gum (30 g) was weighed (M) and transferred into a 100-mL measuring cylinder and the initial volume ($V_0$) recorded. The measuring cylinder was tapped on a bench until a constant volume ($V_f$) of the gum was obtained. The bulk density, tapped density, Carr's index and Hausner ratio were calculated as:

$$\text{Bulk density} = \frac{M}{V_0}$$

$$\text{Tapped density} = \frac{M}{V_f}$$

$$\text{Carr's index} = \frac{(\text{Tapped density} - \text{bulk density})}{\text{Tapped density}} \times 100$$

$$\text{Hausner ratio} = \frac{\text{Tapped density}}{\text{Bulk density}}$$

The fixed height method was also used in the angle of repose determination. The gum was allowed to freely flow through a clamped funnel to form a heap. The height (h) and diameter (d) of the heap were determined. Using the equation below, the angle of repose ($\alpha$) was calculated:

$$\tan(\alpha) = h/(0.5\ d)$$

### *Formulation of tablets*

The wet method of granulation and the technique of doubling the bulk were employed in preparing 12 batches of paracetamol granules. CZG was used as a binder (10% w/v, 15% w/v and 20% w/v) for Batches 1–3 (B1–B3) and as a disintegrant (5% w/w, 7.5% w/w and 10% w/w) for B7–B9. Acacia gum and starch were used as a standard binder and disintegrant, respectively, at the same concentrations stated above for B4–B6 and B10–B12, respectively. All ingredients except talc and magnesium stearate were accurately weighed and mixed together with water as the granulating fluid to form a damp mass. This was then screened with a mesh (2360 $\mu$m) and dried at 60 °C for 2 h in an oven. The dried granules were screened with a 1180 $\mu$m sieve, mixed with magnesium stearate and talc and compressed into tablets using a single punch tableting machine (TDP 5, Herun, China). Sixty (60) tablets were compressed from each batch to cater for all quality control tests. The constituents of the formulated tablets are given in Table 1.

**Table 1:** Constituents (mg) of formulated tablets

| Ingredients | Formulation codes | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | B1 | B2 | B3 | B4 | B5 | B6 | B7 | B8 | B9 | B10 | B11 | B12 |
| Paracetamol | 500 | 500 | 500 | 500 | 500 | 500 | 500 | 500 | 500 | 500 | 500 | 500 |
| *Cinnamomum zeylanicum* gum | 30 | 52 | 64 | – | – | – | 30 | 45 | 60 | – | – | – |
| Starch | 30 | 30 | 30 | 30 | 30 | 30 | – | – | – | 30 | 45 | 60 |
| Acacia | – | – | – | 30 | 52 | 64 | – | – | – | – | – | – |
| Hydroxypropyl methylcellulose | – | – | – | – | – | – | 1.4 | 1.4 | 1.6 | 1.4 | 1.4 | 1.6 |
| Talc | 3.2 | 3.3 | 3.3 | 3.2 | 3.3 | 3.3 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 |
| Magnesium stearate | 6.2 | 6.3 | 6.3 | 6.2 | 6.3 | 6.3 | 6.0 | 6.0 | 6.0 | 6.0 | 6.0 | 6.0 |
| Lactose | 70 | 70 | 70 | 70 | 70 | 70 | 70 | 55 | 40 | 70 | 55 | 40 |

### *Flow properties of granules*

The flow properties of the granules were determined before compression into tablets. The angle of repose was determined using the fixed height method whereas the Carr's index and Hausner ratio were determined using the bulk and tapped densities of the granules.[12,13]

### *Evaluation of tablets*

#### Weight uniformity test

Twenty tablets were randomly selected from each batch and individually weighed to calculate average weight. The average weight and percentage deviation of each tablet were determined as described in the British Pharmacopoeia.[14]

#### Friability test

Ten tablets from each batch were randomly selected and weighed. They were then placed in a friabilator (EF-2W, Roche, Switzerland) regulated at 4 min at a speed of 25 rpm (revolutions per minute). Afterwards, the tablets were removed from the friabilator, dedusted and reweighed.[15] The percentage friability was calculated as:

$$\text{Friability (\%)} = \frac{W_0 - W}{W_0} x100$$

where $W$ is the final weight and $W_0$ is initial weight.

#### Disintegration test

Six tablets from a batch were selected at random and each tablet placed in a tube of the basket-rack assembly in an Erweka disintegration apparatus (ZT 320 series, Germany). The basket rack was placed in a water bath regulated thermostatically at 37±2 °C and observed until all the tablets had disintegrated completely.[12] This procedure was repeated for all batches.

#### Uniformity of dimension and hardness test of tablets

Ten tablets were randomly selected from each batch. Their diameters and thickness were determined using vernier calipers. The hardness of ten randomly selected tablets from each batch was determined with a Veego hardness tester (DIGITAB-SPV, India). The tensile strength was calculated as:

$$\text{Tensile strength} = \frac{2F}{\pi Dt}$$

where D is the diameter, t is the tablet thickness and F is the diametrical break force of the tablet.

#### Uniformity of drug content

Ten tablets from each batch were randomly selected. Each tablet was crushed, dissolved in 50 mL of HCL (0.1 M) and the mixture topped up to 100 mL using 0.1 M HCL. The mixture was filtered and its absorbance determined at 245 nm using a UV spectrophotometer (DU-8800R, Drawell, Shanghai). A standard calibration curve was used to calculate the average drug content for three determinations.

#### In vitro dissolution studies

In vitro drug release of the tablets was determined using the USP II dissolution apparatus (DT6, Erweka, Germany) at 50 revolutions per minute. Six (6) tablets each for all batches were evaluated. A tablet was put in each vessel containing 900 mL of the dissolution medium, HCL (0.1 M), at 37±0.5 °C. A volume of the medium (10 mL) was drawn at time intervals (5, 15, 30, 45 and 60 min) and filtered using Whatman filter paper number 5 with a pore size of 2.5 $\mu$m. The volumes taken were replaced to maintain sink conditions. A UV-visible spectrophotometer (Alpha II, Bruker, Germany) was used to check the absorbance of the filtrates at a wavelength of 245 nm. The average drug release profile for three determinations was carried out using a standard calibration curve.[16] The various batches of tablets which passed the dissolution profile were analysed for their similarity factor (f2) using the equation:

$$\text{Similarity factor (f2)} = 50 x \log [\{1 + (1/n) \sum t=1^{n} (Rt - Tt)^2\}^{-0.5} x 100]$$

where $n$ = dissolution time points, $Rt$ = dissolution value for reference at time $t$, and $Tt$ = dissolution value at time $t$.

### *Statistical analysis*

Statistical analysis was performed using Statistical Package for the Social Sciences (SPSS v.20, IBM). Differences between means were analysed using $t$-tests; $p<0.05$ was considered statistically significant.

## Results and discussion

### *Physicochemical properties of CZG*

The yield obtained from the extractive process was 2.31%. Different solvent systems can affect the percentage yield of gums.[17] According to Vani and Reddy[8] who used an acetone precipitation method, percentage yield obtained for CZG was 1.6%, which is less than the yield obtained in this study (Table 2). Comparing percentage yield obtained from work done by Kolhe et al.[18] using microwave assisted extraction, a higher gum yield of 33.4% was obtained. This could be due to the geographical location of the plant, time or season of harvesting of the plant, and the difference in extraction methods employed.[19] The aqueous method of extraction was used in this study due to its low cost and economic viability. The moisture content of the gum plays a vital role in its flow properties and microbial stability. The moisture content of the extracted gum was 0.63%, which fell within specification (<15% w/w)[20] (Table 2). Gums with moisture content above 15% w/w are likely to stick together, support microbial growth and deteriorate over time. This results in a reduction in shelf life and quality.[21] Therefore, the low moisture content (0.63%) reported for the extracted CZG enhances its suitability as a potential pharmaceutical excipient.

The swelling index of gums contributes to bioadhesion in tablet formulation, which is a determinant of drug release during dissolution. Gums reported to have swelling indices between 31.8% and 180% are said to have good disintegrating properties.[22] The extracted gum exhibited good swelling and water holding capacities (Table 2), which could be attributed to the presence of hydrophilic functional groups. The swelling and water holding capacities of the extracted gum could also be attributed to shorter chain molecules as gums with longer chain polymers are known to have longer hydration time and vice versa.[15,18,19,22] Hence tablets formulated with CZG will potentially exhibit good disintegration and drug release profiles.

The suitability of a natural polymer as an excipient in a formulation can be influenced by its pH. CZG had a neutral pH (Table 2), which indicates its suitability to be incorporated in a pharmaceutical formulation as the stability and physiological activity of the active pharmaceutical ingredient will not be altered.[23]

**Table 2:** Physicochemical characteristics of *Cinnamomum zeylanicum* gum

| Yield (%) | Moisture content (%) | Swelling index (%) | Water-holding capacity (%) | pH |
|---|---|---|---|---|
| 2.31 | 0.63 ± 0.04 | 175 ± 0.14 | 98 ± 0.05 | 6.92 ± 0.08 |

### *Fourier transform infrared spectroscopy analysis*

The spectra of the pure powder of paracetamol, CZG only, and formulated granules revealed the compatibility and stability of the gum and other excipients with the active pharmaceutical ingredients. All principal peaks of the various functional groups for paracetamol (3322.03 $cm^{-1}$, 3159.39 $cm^{-1}$ (hydroxyl group, O-H stretching) and 1561.11 $cm^{-1}$, 1504.98 $cm^{-1}$ (amide II band))[14] were intact in the formulated granules, indicating the absence of interactions between the active ingredients and the excipients (Figure 1).

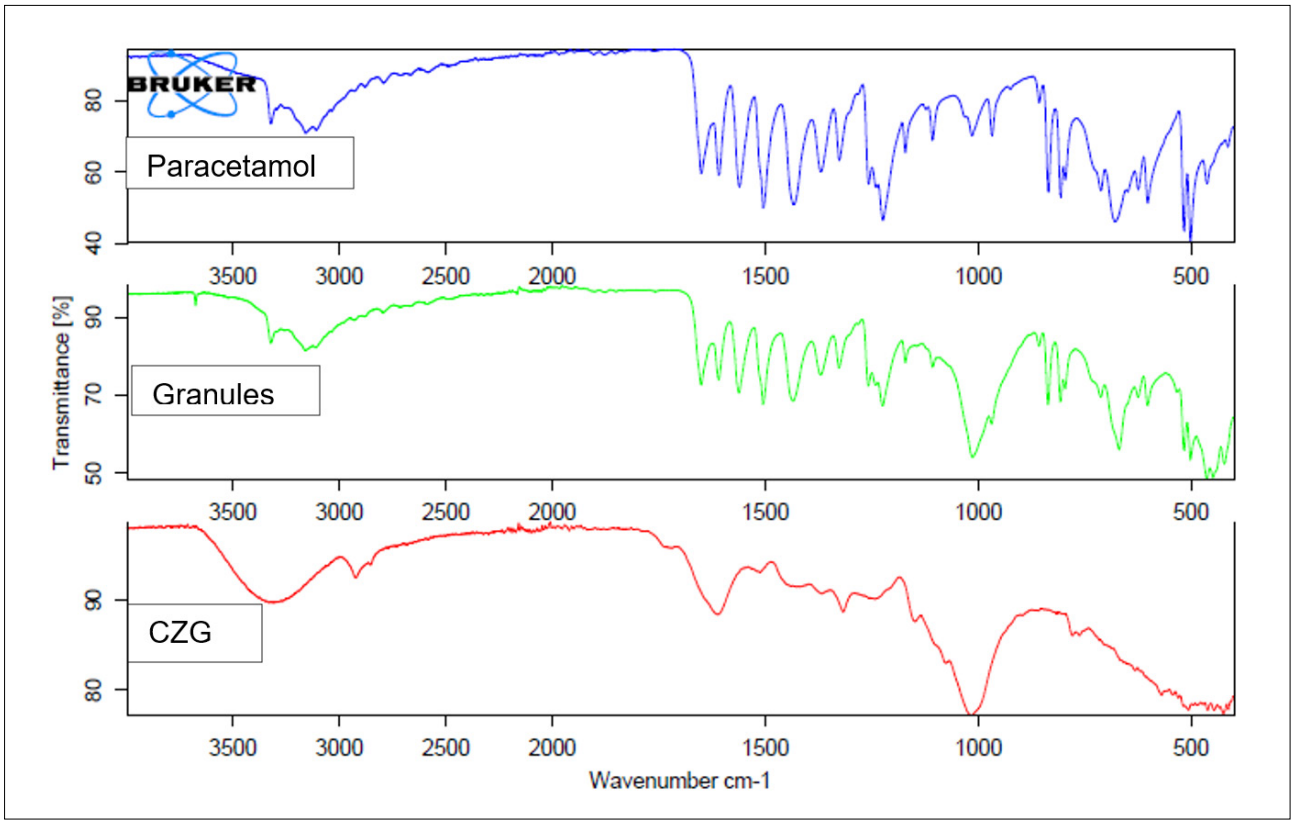



**Figure 1:** Fourier transform infrared spectroscopy spectra of *Cinnamomum zeylanicum* gum (CZG), paracetamol and paracetamol granules with CZG.

### *Evaluation of formulated tablets*

Granules formulated for all batches exhibited good flow with angle of repose, Carr's index and Hausner ratio within the acceptable limits indicating suitability for compression into tablets. This will ensure uniformity in the filling of dies during tablet compression, resulting in weight uniformity, which was evident in the weight uniformity test (Table 3). All formulated tablets exhibited uniformity in diameter and thickness (Table 3), which was within the acceptable limit.[20]

The ability of a tablet to withstand breakage during handling, storage and transportation is affected by its tensile strength and hardness. All formulated tablets had hardness and friability within the specifications of 4–8 kg/f and less than 1%, respectively.[20] Disintegration and drug release of tablets is dependent on the amount and nature of binder and disintegrant used during the granulation process and the compression force applied.[1,5,13] Harder tablets are formed when the concentration of binders is increased with a corresponding extension in the tablet disintegration time.[14,20] On the other hand, the disintegration time of tablets is reduced with an increase in concentration of disintegrants, which was evident in the formulated tablets (Table 3). All the formulated tablets passed the disintegration test (disintegration time less than 15 min) (Table 3). When comparing CZG as a binder and disintegrant to the standard binder (acacia) and standard disintegrant (starch) in immediate release tablets, statistical analysis revealed a significant difference in disintegration time. This indicates that variations in disintegration time between CZG and the standard binder and disintegrant can affect other parameters such as dissolution (Figure 2). There was no significant difference in the friability and hardness profile of formulated tablets for all batches with CZG as binder (B2–B6) and disintegrant (B7–B12) when compared to that of acacia and starch with the exception of B1 and B4 (Figures 3 and 4). Increasing concentration of CZG as binder resulted in tablet with better friability profile, which were comparable to that of the standard binder (Figure 3).

The effectiveness of a dosage form in the gastrointestinal tract and its systemic absorption depends on dissolution and disintegration.[24] At 45 minutes, the amount of drug released should not be less than 70%.[20] A longer disintegration time may result in slower drug release during dissolution which could account for B7 failing the dissolution profile test (Figure 5). B1 and B3 had similar dissolution profiles to that of the standard binder, with only B9 having a similar dissolution profile to that of the standard disintegrant (Table 4). Thus, formulations containing the same active ingredient with similar dissolution profiles are expected to be bioequivalent with similar effects in vivo. CZG can serve as a substitute for acacia and starch for use in the pharmaceutical industry.

## Conclusion

*Cinnamomum zeylanicum* gum exhibited good disintegrating and binding properties at varying concentrations and it can be used as an alternative to starch and acacia as a disintegrant and a binder, respectively, in immediate release tablets.

## Acknowledgement

We are grateful to the technical staff of the Department of Pharmaceutics, Faculty of Pharmacy and Pharmaceutical Sciences, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana, for their support and assistance.

## Competing interests

We have no competing interests to declare.

## Authors' contributions

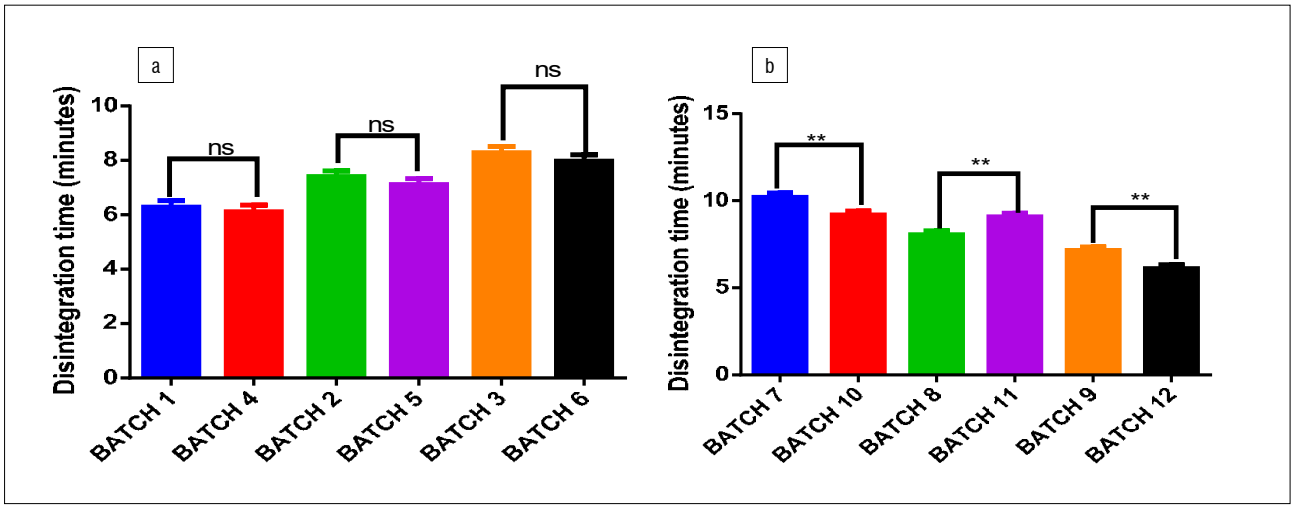
L.C.S.: Methodology, data collection, data analysis, writing – initial draft. P.E.: Conceptualisation, methodology, data collection, data analysis, writing – initial draft, writing – revisions. F.W.A.O.: Methodology, data analysis, writing – revisions. M.E.B.-G.: Conceptualisation, project leadership, project management, methodology, data analysis, writing – revisions. M.T.B. and K.O.-K.: Conceptualisation, project management, methodology, writing – initial draft, writing – revisions.

**Table 3:** Properties of *Cinnamomum zeylanicum* gum (CZG), formulated granules and tablets

| Parameters | CZG | B1 | B2 | B3 | B4 | B5 | B6 | B7 | B8 | B9 | B10 | B11 | B12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Hausner ratio | 1.30±0.03 | 1.17±0.02 | 1.16±0.03 | 1.11±0.02 | 1.18±0.03 | 1.18±0.05 | 1.14±0.04 | 1.09±0.04 | 1.20±0.05 | 1.15±0.04 | 1.16±0.03 | 1.25±0.02 | 1.18±0.03 |
| Carr's index | 23.01±0.02 | 14.380±0.04 | 14.070±0.02 | 10.62±0.03 | 15.38±0.05 | 15.75±0.02 | 12.630±0.04 | 8.40±0.04 | 17.09±0.02 | 13.710±0.03 | 14.080±0.05 | 20.540±0.02 | 15.95±0.03 |
| Angle of repose ($\alpha$) | 30.69±0.02 | 29.510±0.02 | 27.120±0.03 | 28.17±0.02 | 21.48±0.04 | 26.28±0.03 | 30.960±0.02 | 27.34±0.03 | 27.50±0.02 | 26.950±0.04 | 25.160±0.02 | 30.830±0.03 | 28.23±0.02 |
| Tablet diameter (mm) | – | 12.56±0.03 | 12.58±0.05 | 12.61±0.10 | 12.58±0.06 | 12.51±0.02 | 12.39±0.22 | 12.57±0.07 | 12.55±0.03 | 12.55±0.01 | 12.50±0.11 | 12.56±0.04 | 12.47±0.08 |
| Tablet thickness (mm) | – | 3.21±0.04 | 3.29±0.10 | 3.52±0.12 | 3.56±0.07 | 3.34±0.15 | 3.37±0.03 | 3.31±0.14 | 3.43±0.08 | 3.47±0.04 | 3.62±0.06 | 3.96±0.14 | 3.69±0.16 |
| Hardness (kg/f) | – | 4.38±0.07 | 4.26±0.02 | 7.16±0.11 | 4.26±0.06 | 4.62±0.08 | 6.62±0.01 | 4.38±0.15 | 4.38±0.03 | 4.34±0.01 | 4.42±0.04 | 4.48±0.11 | 4.60±0.05 |
| Friability (%) | – | 0.01±0.17 | 0.01±0.01 | 0.01±0.13 | 0.16±0.05 | 0.01±0.19 | 0.01±0.04 | 0.07±0.10 | 0.01±0.11 | 0.05±0.02 | 0.12±06 | 0.03±0.01 | 0.04±0.09 |
| Disintegration time (min) | – | 6.17±0.01 | 7.24±0.05 | 8.17±0.03 | 6.07±0.01 | 7.07±0.04 | 7.58±0.02 | 10.13±0.02 | 8.03±0.07 | 7.09±0.03 | 9.12±0.06 | 9.05±0.04 | 6.07±0.02 |
| Weight uniformity (g) | – | 0.62±0.007 | 0.62±0.004 | 0.61±0.002 | 0.62±0.003 | 0.61±0.001 | 0.62±0.004 | 0.61±0.003 | 0.61±0.002 | 0.62±0.001 | 0.62±0.005 | 0.62±0.001 | 0.62±0.004 |

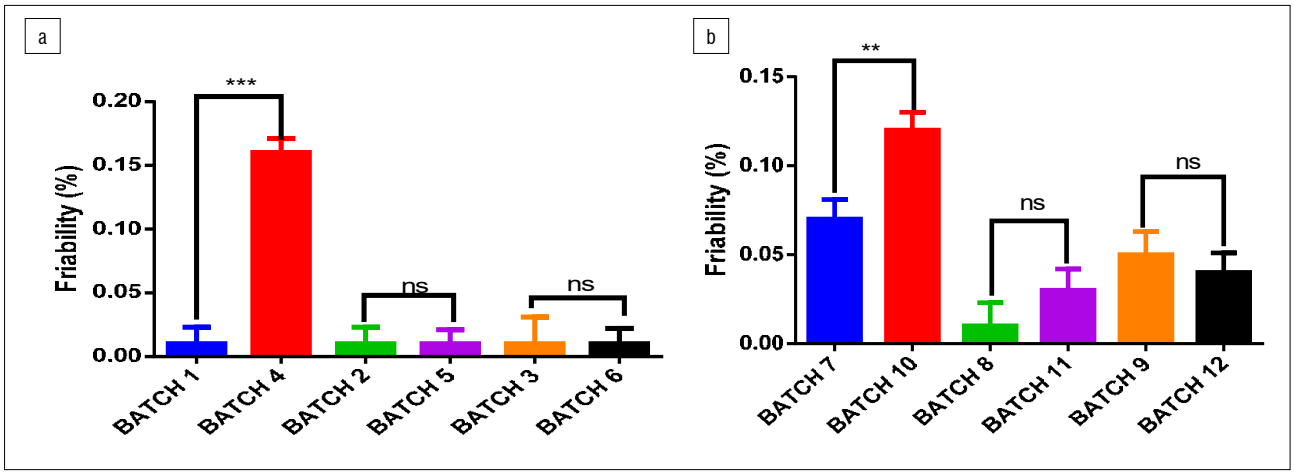
**Table 4:** Similarity factor analysis of formulated *Cinnamomum zeylanicum* gum tablets in comparison with starch or acacia formulations

| Formulation | Similarity factor (*f2*) | Comment |
|---|---|---|
| B1 | 63.50 | Similar |
| B2 | 38.08 | Dissimilar |
| B3 | 60.54 | Similar |
| B7 | 34.51 | Dissimilar |
| B8 | 33.24 | Dissimilar |
| B9 | 53.10 | Similar |



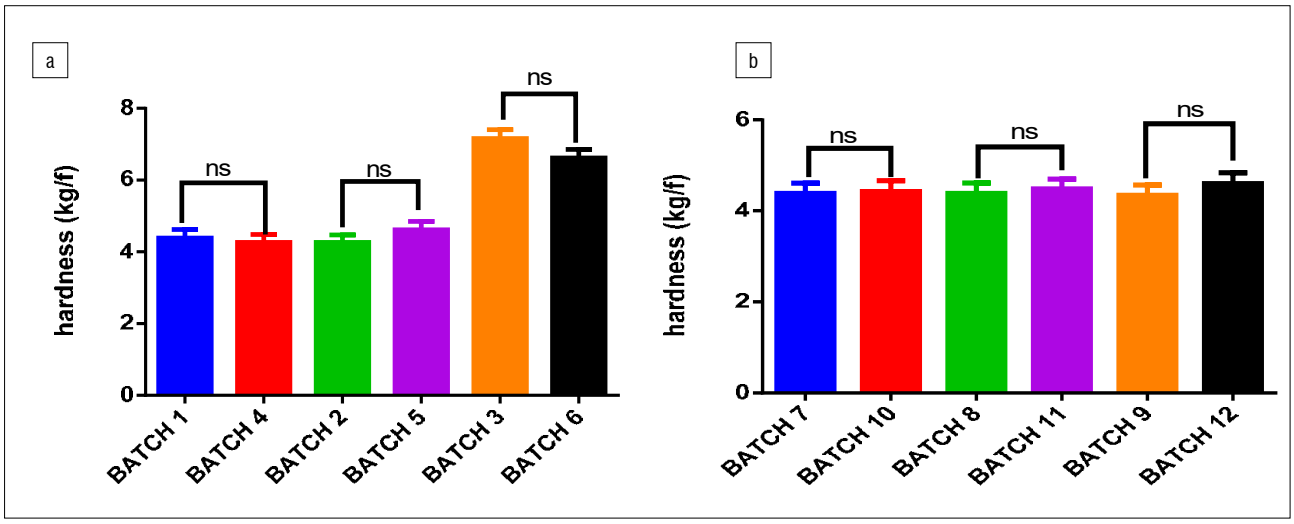

**p≤0.01; ns, not significant; Student's two-tailed t-test

**Figure 2:** Statistical analysis of the disintegration time of paracetamol tablets with (a) *Cinnamomum zeylanicum* gum (B1–B3) or acacia gum (B4–B6) as binders and (b) *Cinnamomum zeylanicum* gum (B7–B9) or starch (B10–B12) as disintegrants.




***p≤0.001; ns, not significant; Student's two-tailed t-test

**Figure 3:** Statistical analysis of the friability index of paracetamol tablets with (a) *Cinnamomum zeylanicum* gum (B1–B3) or acacia gum (B4–B6) as binders and (b) *Cinnamomum zeylanicum* gum (B7–B9) or starch (B10–B12) as disintegrants.




**p≤0.01; ns, not significant; Student's two-tailed t-test

**Figure 4:** Statistical analysis of the hardness of paracetamol tablets with (a) *Cinnamomum zeylanicum* gum (B1–B3) or acacia gum (B4–B6) as binders and (b) *Cinnamomum zeylanicum* gum (B7–B9) or starch (B10–B12) as disintegrants.

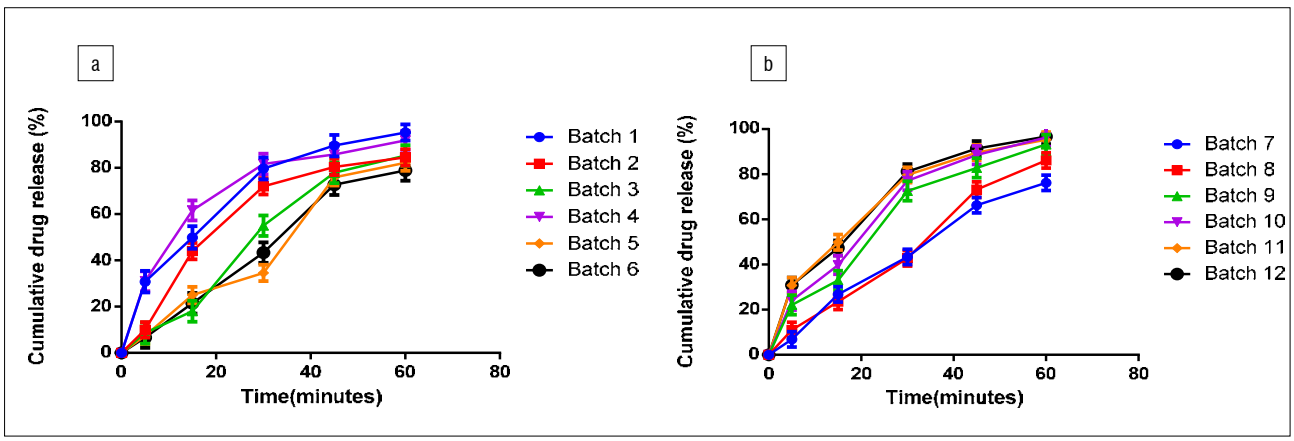



**Figure 5:** Dissolution profiles of (a) Batches 1–6 and (b) Batches 7–12.

## References

1. Chaudhari SP, Patil PS. Pharmaceutical excipients: A review. Int J Adv Pharm Biol Chem. 2012;1(1):21–34.
2. Debnath S, Yadav CN, Nowjiya N, Prabhavathi M, SaiKumar A, Krishna PS, et al. A review on natural binders used in pharmacy. Asian J Pharm Res. 2019;9(1):55–60. https://doi.10.5958/2231-5691.2019.00009.1
3. Patel MT, Patel JK, Upadhyay UM. Assessment of various pharmaceutical excipient properties of natural *Moringa oleifera* gum mucoadhesion, disintegration, binder. Int J Pharm Life Sci. 2012;3(7):1–16.
4. Mistry AK, Nagda CD, Nagda DC, Dixit BC, Dixit RB. Formulation and in vitro evaluation of ofloxacin tablets using natural gums as binders. Sci Pharm. 2014;82(2):441–448. https://doi.org/10.3797/scipharm.1401-14
5. Hussain A, Qureshi F, Abbas N, Arshad MS, Ali E. An evaluation of the binding strength of okra gum and the drug release characteristics of tablets prepared from it. Pharmaceutics. 2017;9(2):20–29. https://doi.org/10.3390/pharmaceutics9020020
6. Entsie P, Owusu F, Boakye-Gyasi ME, Osei YA, Adu F, Bayor MT. Formulation of *Cinnamomum zeylanicum* immediate release tablets as an anthelmintic. Int J Pharm Sci Res. 2021;12(5):2835–2841. https://doi:10.13040/IJPSR.0975-8232.12(5).2835-41
7. Ranasinghe P, Galappaththy P. Health benefits of Ceylon cinnamon. Ceylon Med J. 2016;61(1):1–5. http://doi.org/10.4038/cmj.v61i1.8251
8. Vani YB, Reddy CSP. Formulation and in vitro evaluation of piroxicam emulgel. Int J Pharm Sci Drug Res. 2018;10(4):213–216. https://doi.org/10.25004/ijpsdr.2018.100402
9. Shabana MD. A review on the quality control analysis of oral dosage form: Quality. Res Rev Pharm Sciences. 2016;5(2):108–114.
10. Adjei FK, Osei YA, Kuntworbe N, Ofori-Kwakye K. Evaluation of the disintegrant properties of native starches of five new cassava varieties in paracetamol tablet formulations. J Pharm. 2017;2017, Art. #2326912. https://doi.org/10.1155/2017/2326912
11. Muhammad DRA, Sedaghat Doost A, Gupta V, Bin Sintang MD, Van de Walle D, Van der Meeren P, et al. Stability and functionality of xanthan gum–shellac nanoparticles for the encapsulation of cinnamon bark extract. Food Hyd. 2020;100, Art. #105377. https://doi.org/10.1016/j.foodhyd.2019.105377
12. Alsaifi A, Alyahawi A. Quality assessment of different brands of paracetamol. Uni J Pharm Res. 2018;3(4):42–46. https://doi.org/10.22270/ujpr.v3i4.182
13. Aulton M, Taylor KM. Aulton's pharmaceutics: The design and manufacture of medicines. Edinburgh: Elsevier; 2017.
14. British pharmacopoeia. Vol. 1. London: Medicines and Healthcare products Regulatory Agency; 2018.
15. Zaharuddin ND, Noordin MI, Kadivar A. The use of *Hibiscus esculentus* (okra) gum in sustaining the release of propanolol hydrochloride in a solid oral dosage form. BioMed Res Int. 2014;2014, Art. # 735891. https://doi.org/10.1155/2014/735891
16. Sharma N, Singh S. Central composite designed ezetimibe solid dispersion for dissolution enhancement: Synthesis and in vitro evaluation. Ther Deliv. 2019;10(10):643–658. https://doi.org/10.4155/tde-2019-0063
17. Akdowa EP, Boudjeko T, Woguia AL, Njintang-yanou N, Gaiani C, Scher J, et al. Optimization of variables for aqueous extraction of gum from *Grewia mollis* powder. J Poly. 2014;2014, Art. #926850. https://doi.org/10.1155/2014/926850
18. Kolhe S, Kasar T, Dhole SN, Upadhye M. Extraction of mucilage and its comparative evaluation as a binder. Am J Adv Drug Deliv. 2014;1(2):1–14.
19. Olayemi OJ, Mahmud HS, Apeji Y. Effect of concentration on the release property of *Khaya senegalensis* gum in chloroquine phosphate tablet formulation. Int J Appl Pharm. 2010;2(3):22–26.
20. The United States pharmacopeia. National formulary. Rockville, MD: United States Pharmacopeial Convention; 2015. Paracetamol; p. 2495.
21. Viljoen JM, Steenekamp JH, Marais AF, Kotzé AF. Effect of moisture content, temperature and exposure time on the physical stability of chitosan powder and tablets. Drug Dev Ind Pharm. 2014;40(6):730–742. https://doi.org/10.3109/03639045.2013.782501
22. Bhatta R, Hossain MS, Banik S, Rahman Moghal MM, Rashid MMO, Akter M. Swelling and mucoadhesive behavior with drug release characteristics of gastoretentive drug delivery system based on a combination of natural gum and semi-synthetic polymers. Mar Pharm J. 2018;22(2):286–298. https://doi.org/10.12991/mpj.2018.66
23. Adeyanu O, Lajide L. Physicochemical and binding properties of oxidized *Anacardium occidentale* Linn exudate gum in paracetamol tablet formulations. Afr J Nat Sci. 2011;14:55–59.
24. Nayak AK. Comparative in vitro dissolution assessment of some commercially available paracetamol tablets. Int J Pharm Sci Rev Res. 2010;2(1):29–30.

**AUTHOR:**
Pabalala M. Mthembi[1,2]
Ellen M. Mwenesongole[1,3]
Michael D. Cole[4]

**AFFILIATIONS:**
[1]Department of Genetics, University of the Free State, Bloemfontein, South Africa
[2]South African Police Service Forensic Science Laboratory – Chemistry Section, Pretoria, South Africa
[3]Department of Chemical and Forensic Sciences, Botswana International University of Science and Technology, Palapye, Botswana
[4]Faculty of Science and Engineering, Anglia Ruskin University, Cambridge, United Kingdom

**CORRESPONDENCE TO:**
Pabalala Mthembi

**EMAIL:**
mthembim@saps.gov.za

**DATES:**
**Received:** 07 Aug. 2020
**Revised:** 15 Apr. 2021
**Accepted:** 04 July 2021
**Published:** 29 Nov. 2021

**HOW TO CITE:**
Mthembi PM, Mwenesongole EM, Cole MD. A validated method for the analysis and profiling of 'nyaope' using gas chromatography – mass spectrometry. S Afr J Sci. 2021;117(11/12), Art. #8738. https://doi.org/10.17159/sajs.2021/8738

**ARTICLE INCLUDES:**
☒ Peer review
☐ Supplementary material

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

**EDITOR:**
Priscilla Baker

**KEYWORDS:**
heroin, cannabis, nyaope, GC–MS, sample comparison

**FUNDING:**
University of the Free State

# A validated method for the analysis and profiling of 'nyaope' using gas chromatography – mass spectrometry

Nyaope, a Tswana word for a mixture or 'mish-mash', describes a drug cocktail consisting of heroin, cannabis, and on occasion other controlled substances and warfarin. It is highly addictive with extremely unpleasant side effects caused by withdrawal from the drug. It is a problem drug especially in townships in South Africa. However, its prevalence in neighbouring southern African states and further afield is not yet known. There is currently no validated method for the analysis and comparison of nyaope. We describe a validated method for the gas chromatography – mass spectrometry analysis of nyaope so that within-batch and between-batch comparisons of nyaope can successfully be made for the first time. The validated method managed an accuracy within the range 80–120%, the precision was less than 20% for all analytes and managed linearity with $R^2 \geq 0.99$. The detection limits for diamorphine, efavirenz, nevirapine and $\Delta^9$-tetrahydrocannabinol were 14.2, 18.6, 18.7 and 9.94 pg on column, respectively, and the limits of quantitation were 43.1, 56.3, 56.6 and 30.1 pg on column, respectively. The simulated and casework samples were successfully discriminated into original batches using the identified nyaope components, the unsupervised chemometric methods principal component analysis and hierarchical clustering, as well as chromatographic profiles.

**Significance:**

- A validated method for the analysis and comparison of nyaope allows for data exchange between law enforcement agencies in South Africa and, provided the appropriate quality control measures are in place, between South Africa, neighbouring states and countries further afield. Additionally, public health measures can be put in place now that it is possible to use a validated method to determine the contents of nyaope.

## Introduction

Despite the first reports being in the early 2000s[1,2], there remains a dearth of analytical chemistry and forensic science literature on the drug 'nyaope'. Nyaope, the name of which is derived from the Tswana word for mixture or 'mish-mash'[2], is reported to be a mixture of powdered heroin, herbal cannabis and prescription antiretrovirals. It has also been reported to contain methamphetamine and warfarin, although some of these claims are disputed. An unpublished study reports a much wider range of drugs found in nyaope samples.[3] In parts of South Africa, it is known as 'Woonga' (spelled 'Wunga' in Zulu).[2] It has no single composition and mixtures follow trends[4], although the predominant drug by mass is cannabis, followed by heroin. It is typically smoked after being mixed with tobacco.

Nyaope contains substances that have been controlled by the *South African Drug and Drug Trafficking Act* since1992. The name nyaope itself is a street name and therefore cannot be listed in the Act. It is inexpensive, with a single dose costing ZAR20.00– 30.00 (circa USD2), and is widely used by the coloured and black communities in South African townships, by addicts who are sometimes as young as 12 or 13 years old.[4] In a conversation with a colleague from the Portuguese criminal police (Alegre J 2019, oral communication, May 08), it was established that, although the drug is widely reported in South Africa, nothing is known of its distribution in neighbouring Mozambique, Zimbabwe, Botswana and Namibia, although there are anecdotal reports of its use in Angola. Because of the severe form that the addiction to nyaope can take, as well as the intensity and nature of the withdrawal symptoms, the social cost of this drug is enormous[5] and addicts, including children, adopt chaotic lifestyles, such as dropping out of school and engaging in criminal activity including theft and prostitution, to fund their drug taking[4].

The health risks associated with using nyaope are not well investigated but reports include restriction of growth and development in neonates[6] and infective endocarditis[7] which had been misdiagnosed elsewhere as pulmonary tuberculosis or pneumonia[8]. The vast majority of users are HIV positive.[7] Other problems associated with nyaope use include damaged and infected veins, damaged heart valves, tissue infections, liver failure, kidney disease and lung problems.[9] It therefore represents a significant public health risk.

In South Africa, at present, the criminal law punishes drug-related offences by a fine or imprisonment. The scale of the nyaope problem is difficult to quantify. At present, there are few representative surveys on drug use and abuse in South Africa. A recent policy brief summarises the rise in trafficking and abuse of heroin in South Africa.[10] However, there is a growing thesis that to tackle the drug problem in South Africa, including that associated with nyaope, a number of approaches need to be taken.[4] These approaches include (1) punishment of those manufacturing, trafficking and distributing the drug. Additionally, a forensic care process is proposed towards addressing the nyaope problem where (2) drug users are properly catered for by rehabilitation schemes, (3) the social circumstances of the drug users are changed and (4) a Public Health Awareness Scheme is implemented.[4] Of course, these approaches could be applied to many drug problems across the globe.

In order that (1) and (4) can be achieved and supported, it is necessary to prove the identity of the drug as nyaope and then to identify and quantify the drug contents. In order to achieve this identification, a validated analytical method for the analysis of the contents of and comparison of nyaope is required – to date no such method has been developed. This is further exacerbated by the fact that the usual methods for cannabis analysis cause heroin to break down, and, conversely, those used for heroin cause cannabis to break down. Our previous studies have identified that two forensic science questions need to be identified before police operations involving nyaope commence[11]: how should nyaope be stored post-seizure[11], and how should the drug be prepared prior to analysis by gas chromatography – mass spectrometry (GC–MS)[12]. In this paper, we present, for the first time, a validated analytical method for the identification of nyaope and quantification of the drug components, which addresses the analysis of heroin and cannabis when contained in the same drug sample. We further demonstrate that, with correct collection, storage and sample preparation, it is possible to compare nyaope samples, identify those which are related and discriminate between those that are not, using, for the first time, chemometric clustering techniques. How this method will assist law enforcement and public health officials in South Africa, and further afield, is also discussed.

## Materials and methods

### Chemicals

Tertiary butyl alcohol (t-BuOH) was purchased from Merck (Darmstadt, Germany), tetracosane 99% was purchased from Sigma-Aldrich (St Louis, MO, USA), isopropanol (i-PrOH)-AR grade was purchased from Associated Chemical Enterprise (Johannesburg, South Africa). Representative compounds, identified in casework samples of nyaope by the South African Police Service (SAPS), were used to validate the GC–MS method. Certified reference standards of $\Delta^9$-tetrahydrocannabinol ($\Delta^9$-THC) and diamorphine (both 1 mg/mL) were purchased from Cerilliant-Sigma Aldrich (Austin, TX, USA). Caffeine and phenacetin were purchased from the US-Pharmacopeia (Rockville, MD, USA) as USP powder reference standards while efavirenz and nevirapine were purchased from WHO International Chemical Reference Substances (Strasbourg, France) as ICRS powder reference substances.

### Preparation of internal standards

The internal standard solution, tetracosane ($C_{24}$), was prepared at a final concentration of 0.02 mg/mL in tertiary butyl alcohol. Tertiary butyl alcohol has previously been shown to be the solvent of choice for presenting nyaope extracts to the GC–MS.[12] The internal standard solution was used to dilute the certified reference standards, and other samples, before GC–MS analysis.

### Preparation of calibration standards

Stock solutions (1 mL at 1 mg/mL) of $\Delta^9$-THC in methanol and diamorphine in acetonitrile were placed in an amber GC–MS vial, evaporated to dryness under nitrogen and then re-dissolved in 1 mL of the internal standard solution to give stock solutions of 1 mg/mL. Phenacetin, caffeine, efavirenz and nevirapine were dissolved at concentrations of 1.03, 1.00, 0.998 and 1.05 mg/mL, respectively, in the internal standard solution. From these, 14 standards in the concentration range 0–1.0 mg/mL, at notional concentrations of 0, 0.001, 0.0025, 0.005, 0.0075, 0.01, 0.025, 0.05, 0.075, 0.1, 0.25, 0.5, 0.75 and 1.0 mg/mL were prepared.

### Instrumentation

GC–MS analysis was carried out using an Agilent Technologies system (Chemetrix, RSA) consisting of a gas chromatograph, Agilent 7890A, and mass selective detector (Agilent 5975 CVL MSD) with an auto sampler 7683 B series (1 $\mu$L injection). Chromatographic separation was performed using a computer-controlled autosampler with a fused-silica capillary column HP-5MS (30 m x 0.25 mm i.d., film thickness 0.25 $\mu$m; J&W Scientific, Folsom, CA, USA). Splitless injection was used at 280 °C. The GC oven temperature programme consisted of an initial temperature of 100 °C for 0.4 min, raised to 290 °C at a rate of 60 °C/min, held at 290 °C for 2.4 min then raised to reach 316 °C at 60 °C/min and held for 3 min. The total run time was 9.4 min. High-purity helium (99.9995%) was used as the carrier gas, at a flow rate of 1 mL/min. The MS parameters used were as follows: the interface temperature was 280 °C, the inlet temperature 250 °C, the ion-source temperature 230 °C, electron ionisation was achieved with a source voltage of 70 eV and the mass spectrometer (quadrupole) was used in scan mode. The spectra were recorded in the scan range (m/z) 35 to 550 amu, at a scan rate of 1 scan/s.

### Method validation

The method was validated by determining the precision of the retention index of each compound, the linearity of detector response, the limit of detection and of quantitation, repeatability and the reproducibility of the measurements.[13,14] The precision of the retention index was obtained for each compound (phenacetin, caffeine, efavirenz, nevirapine, $\Delta^9$-THC and diamorphine) by calculating the mean, standard deviation and relative standard deviation of the retention index, relative to tetracosane, for 10 replicate analyses.

Linearity of the detector response to the exemplar drugs was determined by preparation of calibration curves for samples in the concentration range 0.00–1.00 mg/mL. The regression equations for detector response relative to the internal standard, the value of $R^2$ and residual plot analysis were used to confirm linearity of detector response.

Limit of detection (LOD) and limit of quantitation (LOQ) were determined by the calibration curve slope using reference sample solutions with concentrations in the vicinity of the LOD[15], namely 0.000, 0.001, 0.0025, 0.005, 0.01 and 0.05 mg/mL and the equations:

$$LOD = \frac{3.3 \times \sigma \times C_{IS}}{S} \quad \text{Equation 1}$$

$$LOQ = \frac{10 \times \sigma \times C_{IS}}{S} \quad \text{Equation 2}$$

where $\sigma$=standard error of the measured response, S=slope of the regression line and $C_{IS}$=concentration of the internal standard = 0.018 mg/mL

To measure the accuracy of the method (closeness to true concentration), 10 replicate measurements of standards of known concentrations were made, the experimental concentrations determined and Equation 3 applied:

$$\%\ Accuracy = \frac{Measured\ concentration}{actual\ concentration} \times 100 \quad \text{Equation 3}$$

Precision is a measure of the closeness of the analytical results obtained from a series of replicate measurements of the same measure under the conditions of the method. Intra-assay precision (repeatability) and inter-assay precision (reproducibility) were assessed using drug standard mixtures of phenacetin, caffeine, efavirenz, nevirapine, $\Delta^9$-THC and diamorphine at three concentration levels (0.01, 0.1 and 1.00 mg/mL). Repeatability was assessed by making 10 replicate analyses of the drug standards at three concentration levels and calculating the mean, standard deviation and relative standard deviation of the relative response to the internal standard.

Reproducibility was assessed by making five replicate analyses of the drug standards over five consecutive days at the three concentration levels, and calculating both within group (W) and between group (B) precision using one-way ANOVA (Group = Day)[16]:

$$\%RSD_w = \frac{\sqrt{MSW}}{x} \times 100 \quad \text{Equation 4}$$

$$\%RSD_B = \frac{\sqrt{(MSB - MSW)/n}}{x} \times 100 \quad \text{Equation 5}$$

If MSB < MSW, set $\%RSD_B = 0$

where RSD = relative standard deviation, $X$ = grand mean of all observations, $n$ = number of observations in group, MSW = mean of squares within group, MSB = mean of squares between groups.

### *Nyaope sample profiling and comparison*

In order to investigate the validity of the method for nyaope sample identification and comparison, both simulated samples and casework samples of nyaope were analysed.

Street cannabis and heroin samples seized by the SAPS were used to prepare simulated nyaope samples. Three blind simulated samples were prepared by mixing a heroin street sample, a cannabis street sample, efavirenz tablets and nevirapine tablets, in different combinations and proportions, to mimic as closely as possible a typical street nyaope sample. The three mixtures were homogenised by grinding using a mortar and pestle. The samples were then further divided into six blind sub-samples each to give a total of 18 samples marked S1–S18. Homogenised samples which had a mass ranging between 12 mg and 14 mg were mixed with 3 mL of the internal standard solution (0.02 mg/mL tetracosane in tertiary butyl alcohol) in a 20-mL head space vial. The mixture was sonicated for 15 min[12,17,18], filtered into amber GC–MS vials and analysed in triplicate. Each of the extracts of the simulated samples S1–S18 was analysed at 0, 24, 48 and 72 h in order to confirm the stability of the extract once prepared.[13] Additionally, chromatograms of members of each of the three groups were compared at the same time intervals to determine whether samples from the same parent batch could be discriminated after these time intervals. Finally, chromatograms from one member of each of the three groups were compared at these time intervals to demonstrate whether it is possible to discriminate between groups.

Five casework samples of nyaope were ground into a fine powder using a mortar and pestle. Sub-samples (circa 12–14 mg) of these street samples were placed in a 20-mL vial and extracted with 3 mL of the internal standard solution prior to analysis. Each of the casework samples was analysed in triplicate by GC–MS at t=0 after extraction. The chromatograms were compared to determine whether it was possible to discriminate between street samples. Each extract was then analysed after 24, 48 and 72 h to demonstrate stability of extracts of such samples. Semi-quantitation was conducted on caffeine, diamorphine and $\Delta^9$-THC for each of the five casework samples.

Two unsupervised chemometric methods – agglomerative hierarchical cluster analysis (HCA) and principal component analysis (PCA) – were performed on both the blind simulated and casework nyaope samples using the XLSTAT statistical and data analysis solution 2019 version. The HCA and PCA analysis were conducted for the samples analysed at 0, 24, 48 and 72 h.

## Results and discussion

### *Compound identification – stability of retention indices*

It is important that retention indices are stable for a given analytical method if drug comparisons are to be made. The retention indices were evaluated over five days for each component. The stability data for representative components of nyaope are given in Table 1. The ANOVA test (single factor) demonstrated ($F_{calc}$ = 0.029, $F_{crit}$ = 4.965) that there was no significant difference between days and that the retention indices were stable.

The relative standard deviation (Table 1) for the retention indices of the compounds caffeine, diamorphine, efavirenz, nevirapine, phenacetin and $\Delta^9$-THC were all below 0.025%, further illustrating the stability of this parameter. Identification of components of nyaope can therefore be made on the basis of retention index and the mass spectrum of each separated compound.

**Table 1:** Stability data for retention indices of representative components of nyaope measured by gas chromatography – mass spectrometry

| Compound | Average | SD | %RSD |
|---|---|---|---|
| Phenacetin | 0.666 | 0.000148 | 0.022 |
| Caffeine | 0.835 | 0.000169 | 0.020 |
| Efavirenz | 0.835 | 0.000169 | 0.020 |
| Tetracosane | 1.000 | 0.000000 | 0.000 |
| Nevirapine | 1.096 | 0.000209 | 0.019 |
| $\Delta^9$-Tetrahydrocannabinol | 1.178 | 0.000137 | 0.012 |
| Diamorphine | 1.361 | 0.000332 | 0.024 |

### *Linearity of detector response*

The detector response to standard compounds was linear over the concentration range investigated. The regression equations and $R^2$-values are given for each of the components of nyaope measured (Table 2). All of the $R^2$-values are above 0.99. This and analysis of residuals demonstrate that the detector response to these drugs is linear.

### *Limits of detection and quantitation*

The detection limits and the limits of quantitation using this analytical method were determined as mass of the free drug on column and are given in Table 2. The detection limits varied between 9.94 pg and 39.1 pg on column and the limit of quantitation between 30.1 pg and 118 pg on column. The method is sufficiently sensitive to both detect these drugs in nyaope and quantitate them in a street sample.

### *Accuracy*

Ten replicate analyses were performed for each drug at each of three known concentrations and the percentage accuracy of the measurement was determined. Accuracy figures were found to lie between 82% and 112%. These lie between the limits of 80% and 120% and are therefore considered accurate[19-21], further validating the method.

**Table 2:** Regression equations, $R^2$-values, limits of detection (LOD) and limits of quantitation (LOQ) for the exemplar drugs in nyaope

| Drug | Regression equation | $R^2$-value | LOD (pg) | LOQ (pg) |
|---|---|---|---|---|
| Caffeine | $y=23.94x+0.3923$ | 0.9975 | 21.0 | 63.6 |
| Diamorphine | $y=31.36x+0.3186$ | 0.9981 | 14.2 | 43.1 |
| Efavirenz | $y=30.92x+0.1483$ | 0.9995 | 18.6 | 56.3 |
| Nevirapine | $y=29.49x+0.1673$ | 0.9987 | 18.7 | 56.6 |
| Phenacetin | $y=28.05x+0.3406$ | 0.9970 | 39.1 | 118 |
| $\Delta^9$-Tetrahydrocannabinol | $y=40.03x+0.499$ | 0.9951 | 9.94 | 30.1 |

### Repeatability and reproducibility

Repeatability was assessed through 10 replicate analyses of the drugs at each of three different concentrations. The relative standard deviations of all analyses were found to lie below or at 15% (with nevirapine just above 15%, 0.01 mg/mL, RSD = 15.78%), demonstrating that the method is repeatable.[22,23]

Reproducibility was assessed through five replicate analyses of the drug standards. The relative standard deviations for both within-group and between group precision were found to lie below 15% (with caffeine and nevirapine just above 15%, 0.01 mg/mL, RSD = 15.01% and 15.19% respectively) demonstrating that the method is reproducible.[14,19,20]

On the basis of the data described above, and the recommendation of the UNODC that cannabis and heroin can be analysed by GC–MS, without derivatisation[18,24], the method was deemed suitable for the analysis of the principal drug types in nyaope.

### Nyaope sample profiling and comparison

Each of the simulated samples was analysed after 0, 24, 48 and 72 h. The average peak area ratio (PAR) was determined for each of the simulated samples. The ANOVA of one of the samples gave $F_{calc} = 0.0106 < F_{crit} = 2.798$, demonstrating that there was no significant difference between the PAR over the 72 h. Retention time data are provided in Table 3. Pooled average response ratios for each of the batches were determined by averaging the PAR at t = 0, 24, 48 and 72 h. ANOVA of the samples from one of the batches using the F-test (single factor) gave $F_{calc} = 0.0268 < F_{crit} = 2.342$, demonstrating that there were no significant differences amongst the PARs for the samples belonging to the same batch over the 72 h autosampler storage.

**Table 3:** Retention time and relative retention time for the compounds identified in the analysis of simulated and casework nyaope samples. Identifications were made on the basis of retention indices and mass spectra of standards and casework samples (data available from corresponding author)

| | Retention time | Relative retention time |
|---|---|---|
| Nicotine | 2.411 | 0.444 |
| Caryophyllene | 2.582 | 0.476 |
| Bulnesol | 3.098 | 0.571 |
| Phenacetin | 3.133 | 0.577 |
| Acetaminophen | 3.194 | 0.589 |
| Caffeine | 3.535 | 0.651 |
| Efavirenz | 4.545 | 0.837 |
| Methaqualone | 4.748 | 0.875 |
| Cocaine | 4.891 | 0.901 |
| Tetrahydrocannabivarin | 5.241 | 0.966 |
| Tetracosane | 5.427 | 1.00 |
| Cannabivarin | 5.573 | 1.027 |
| Cannabichromene | 5.731 | 1.056 |
| Cannabidiol | 5.754 | 1.060 |
| Nevirapine | 5.949 | 1.096 |
| $\Delta^9$-Tetrahydrocannabinol | 6.234 | 1.149 |
| Cannabigerol | 6.414 | 1.182 |
| Acetylcodeine | 6.501 | 1.198 |
| Cannabinol | 6.518 | 1.201 |
| 6-Monoacetylmorphine | 6.571 | 1.211 |
| Diamorphine | 7.020 | 1.293 |
| Nonacosane | 7.517 | 1.385 |
| Vitamin E | 8.686 | 1.601 |

From the data obtained, it can be demonstrated that each compound of interest can be identified, and that their relative proportions do not change over a 72-h period once extracted from nyaope into tertiary butyl alcohol, thus confirming previous work.[12] When each of the six samples for the three different simulated samples was analysed, it was found that the PAR, relating to each batch, could not be separated. It is therefore now possible to relate samples of nyaope to each other when they have come from a once larger parent batch, as demonstrated by the PAR.

When separated by GC–MS, it was possible to differentiate between the three batches of simulated nyaope (Figure 1) on the basis of the chromatographic profile. This method of extraction and analysis allows, for the first time, analysis and comparison of nyaope samples by a forensic science laboratory.

To demonstrate that the method could be applied to casework samples, five samples of nyaope (denoted as 2514202B, 2520902B, 3400002B, 37959902B and 50390902B) were analysed. Figure 2 is a typical total ion chromatograph for one of the casework samples (37959902B). The components identified in the other casework samples are summarised in Table 4. The components – cocaine, diamorphine, methaqualone and $\Delta^9$-THC – were identified on the basis of their retention time and mass spectral data using certified reference material. Caffeine was identified on the basis of retention time and mass spectral data using USP reference standards. Acetaminophen, acetylcodeine, bulnesol, cannabichromene, cannabicoumaronone, cannabidiol, cannabigerol, cannabinol, cannabivarin, codeine, caryophyllene, 6-monoacetylmorphine, nicotine, nonacosane, tetrahydrocannabivarin and vitamin E were identified by comparing the experimental mass spectral data with the NIST mass spectral library NIST 14. From the components identified, it can be seen that it is possible to discriminate between street samples of nyaope and that the method described can be applied to forensic casework.

Semi quantitation of the five casework samples is summarised in Table 5, which shows the concentrations of the components caffeine, diamorphine and $\Delta^9$-THC. The caffeine concentration for samples 2514202B and 2520902B was below the limit of quantitation (63 pg on column). The pooled average concentration for each time interval of the three sub-samples in a batch was used to calculate the average %RSD, shown in Table 5, in order to determine if the concentrations were significantly different. The average %RSD was found to be <10% for all the components (caffeine, diamorphine and $\Delta^9$-THC) in the samples. This indicates that there was no significant difference between the concentrations of a particular component, over a period of 72 h once extracted into tertiary butyl alcohol. This suggests that all the components were stable for the 72 h of autosampler stability, confirming the previous finding.[12]

To demonstrate that the extracts of casework samples are stable for up to 72 h after extraction, a chromatographic analysis of the five casework samples was undertaken. The ANOVA of the PAR of one of the casework samples (37959902B) using F-test (single factor) gave $F_{calc} = 0.0429 < F_{crit} = 3.285$, demonstrating that there were no significant differences between the PAR over the 72 h. ANOVA of the pooled average response ratios gave $F_{calc} = 0.0429 < F_{crit} = 3.285$, demonstrating that there were no significant differences amongst the PARs for the samples belonging to the same group over the 72-h autosampler storage. From these data, it can be concluded that, as with simulated nyaope samples, casework samples of nyaope are stable up to 72 h after the preparation of drug extracts into tertiary butyl alcohol.

### Hierarchical cluster analysis

HCA, using agglomerative clustering and unweighted linkage, was conducted on the average concentrations of caffeine, diamorphine, efavirenz, nevirapine, phenacetin and $\Delta^9$-THC for each of the 18 blind simulated and the 5 casework nyaope samples. The matrices generated for the HCA clustering indicated in Table 6 for the time interval t=72 h, demonstrate that the HCA method was suitable for discriminating the samples into different classes.[25] The matrices demonstrate that there was a maximum distance between Class 2 and Class 3 for the simulated samples and between Class 1 and Class 3 for the casework samples.

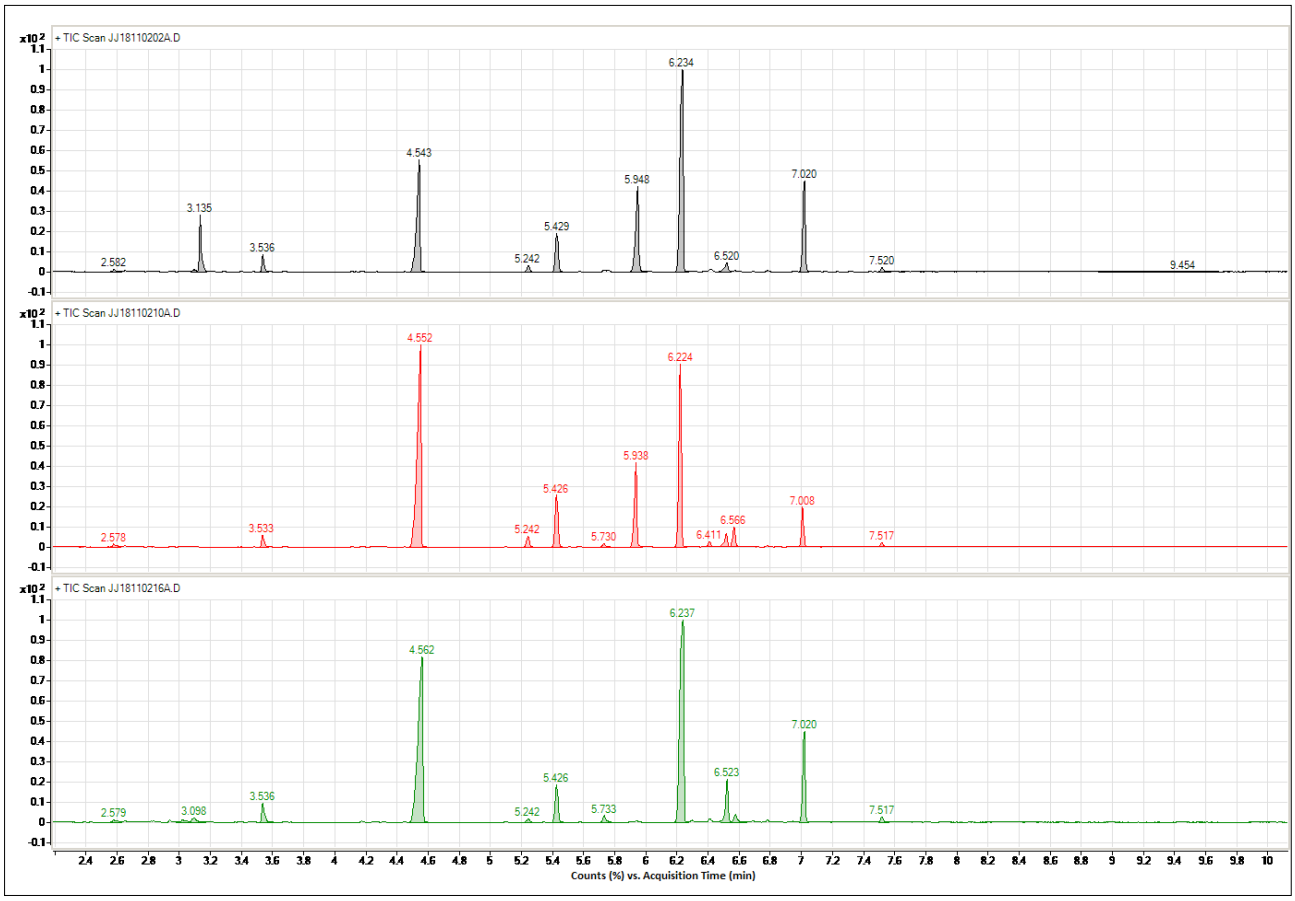



**Figure 1:** Exemplar chromatograms obtained by gas chromatography – mass spectrometry analysis of the three batches of simulated nyaope showing that it is possible to differentiate different batches of the drug, where identifications were (2.582) caryophyllene, (3.135) phenacetin, (3.536) caffeine, (4.543) efavirenz, (5.242) tetrahydrocannabivarin, (5.429) tetracosane IS, (5.733) cannabichromene, (5.948) nevirapine, (6.234) $\Delta^9$-THC, (6.411) cannabigerol, (6.520) cannabinol, (7.020) diamorphine, (7.520) nonacosane, (9.454) unknown in the first chromatogram.

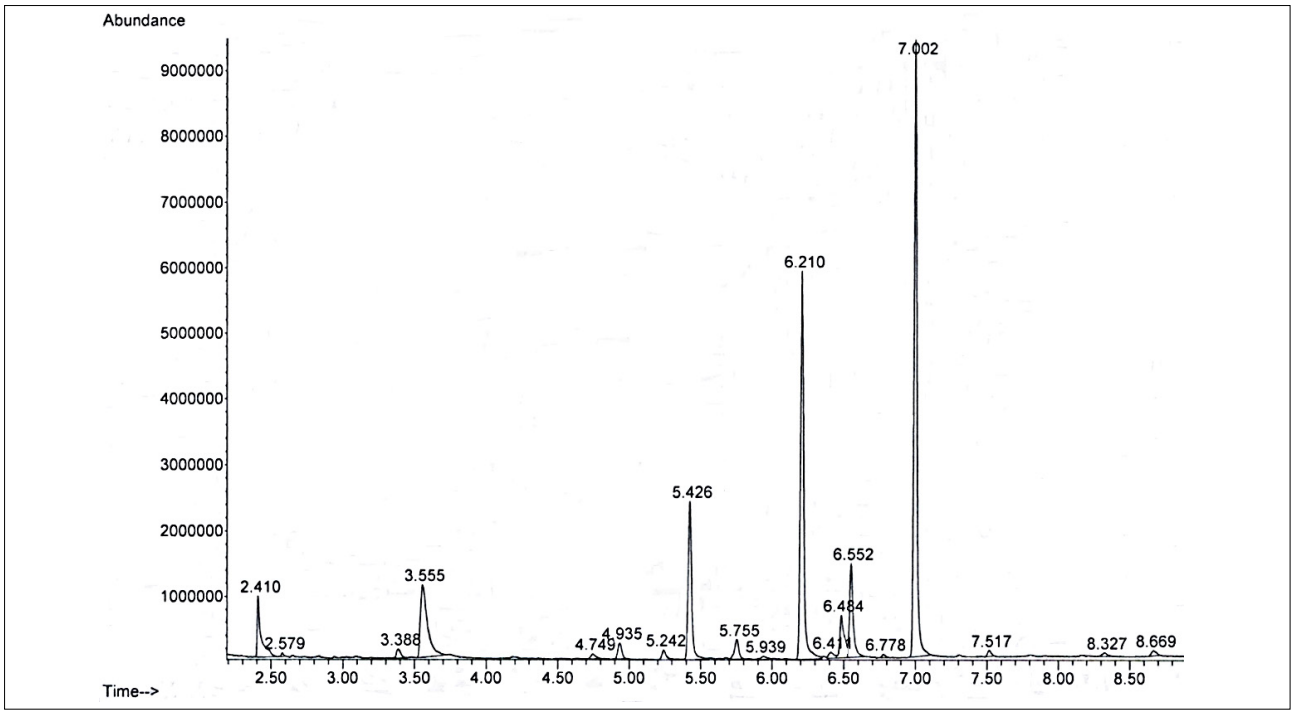



**Figure 2:** A typical total ion chromatograph for one of the casework samples (37959902B): (2.410) nicotine, (2.579) caryophyllene, (3.388) unknown, (3.555) caffeine, (4.749) methaqualone, (4.935) unknown, (5.242) tetrahydrocannabivarin, (5.426) tetracosane IS, (5.765) cannabidiol, (5.939) codeine, (6.210) $\Delta^9$-THC, (6.411) cannabigerol, (6.484) acetylcodeine, (6.552) 6-monoacetylmorphine, (7.002) diamorphine, (7.517) nonacosane, (8.327) unknown, (8.669) vitamin E.

**Table 4:** Components identified in the five casework samples

| Sample component | 2514202B | 2520902B | 3400002B | 37959902B | 50390902B |
|---|---|---|---|---|---|
| Bulnesol | × | √ | × | × | × |
| Acetaminophen | √ | × | × | × | × |
| Acetylcodeine | × | × | √ | √ | × |
| Caffeine | √ | √ | √ | √ | √ |
| Cannabichromene | √ | √ | √ | × | × |
| Cannabicoumaronone | × | √ | √ | × | × |
| Cannabidiol | × | × | × | √ | × |
| Cannabigerol | √ | √ | × | √ | × |
| Cannabinol | √ | √ | × | × | × |
| Cannabivarin | × | √ | × | × | × |
| Cocaine | × | × | √ | × | × |
| Codeine | × | × | × | √ | × |
| Diamorphine | √ | × | √ | √ | √ |
| Methaqualone | × | √ | × | √ | √ |
| 6-monoacetylmorphine | √ | × | √ | √ | √ |
| Nicotine | × | √ | × | √ | √ |
| Nonacosane | √ | √ | × | √ | × |
| Tetrahydrocannabivarin | √ | √ | √ | √ | × |
| $\Delta^9$-Tetrahydrocannabinol | √ | √ | √ | √ | √ |
| Vitamin E | × | × | × | √ | × |

√ = *identified*
× = *not identified*

**Table 5:** Concentration of actual street nyaope samples (mg/mL x 1000)

| Sample | Caffeine | | | | %RSD | Diamorphine | | | | %RSD | $\Delta^9$-Tetrahydrocannabinol | | | | %RSD |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0 h | 24 h | 48 h | 72 h | | 0 h | 24 h | 48 h | 72 h | | 0 h | 24 h | 48 h | 72 h | |
| 2514202B | < QL | < QL | < QL | < QL | **< QL** | 0.21 | 0.20 | 0.21 | 0.21 | **4.00** | 5.48 | 4.97 | 5.21 | 5.25 | **4.00** |
| 2520902B | < QL | < QL | < QL | < QL | **< QL** | nd | nd | nd | nd | **nd** | 1.44 | 1.39 | 1.32 | 1.32 | **4.34** |
| 3400002B | 0.50 | 0.50 | 0.49 | 0.49 | **1.47** | 2.53 | 2.42 | 2.44 | 2.38 | **2.52** | 0.32 | 0.33 | 0.31 | 0.30 | **3.92** |
| 37959902B | 0.44 | 0.42 | 0.42 | 0.42 | **2.19** | 2.10 | 2.08 | 2.11 | 2.03 | **1.67** | 0.87 | 0.85 | 0.78 | 0.84 | **4.70** |
| 50390902B | 0.82 | 0.80 | 0.81 | 0.81 | **0.94** | 2.87 | 2.82 | 2.82 | 2.79 | **1.10** | 0.11 | 0.10 | 0.10 | 0.10 | **6.36** |

*<QL, below quantification limit; nd, not detected*

**Table 6:** Matrices showing distances between central objects at t=72 h for the simulated nyaope samples S2, S3 and S18 and casework samples A2, B3 and C1

| Blind simulated samples | | | | Case work samples | | | |
|---|---|---|---|---|---|---|---|
| **Class** | **1 (S2)** | **2 (S3)** | **3 (S18)** | **Class** | **1 (A2)** | **2 (B3)** | **3 (C1)** |
| **1 (S2)** | 0 | 2.92 | 2.34 | **1 (A2)** | 0 | 3.72 | 5.22 |
| **2 (S3)** | 2.92 | 0 | 3.63 | **2 (B3)** | 3.72 | 0 | 2.62 |
| **3 (S18)** | 2.34 | 3.63 | 0 | **3 (C1)** | 5.22 | 2.62 | 0 |

They further demonstrate that there was a minimum distance between Class 1 and Class 3 for the simulated samples and between Class 2 and Class 3 for the casework samples.

The results of the HCA performed on both the simulated and casework nyaope samples is also indicated in the dendrograms in Figure 3 for the time interval t=72 h. The results demonstrate that HCA successfully discriminated both the simulated and casework samples into three and five different batches, respectively. The HCA further demonstrated that both simulated and casework nyaope samples can still be discriminated even after 72 h of autosampler storage.

### *Principal component analysis*

The correlation matrices for the PCA analysis of both simulated and casework nyaope samples (Table 7) indicate that all values are different from zero at a significance level of $\alpha=0.95$ (two-tailed), which indicates that there is a linear correlation between the variables.[26] The transpose of these correlation matrices are identical to the matrices themselves. As a result, their product would yield identity matrices, which demonstrates that the PCA is orthogonal.[26] The PCA indicates that there are three principal components (F1, F2 and F3) for the simulated samples and one principal component (F1) for the casework samples that explain the variability of the variables.

The eigenvalues of these principal components were greater than 1.00. The total variability (%) of the principal components (F1 and F2) for the simulated samples was more than the minimum 70%, while the principal component F1 alone accounted for more than 70% of the variability for the casework sample. This further indicates that the two principal components, F1 and F2, for the simulated samples, and the single component F1 for the casework sample, are sufficient to explain the variability of the data set.[27,28] As indicated in Figure 4, the PCA discriminated the simulated samples into three different batches and the casework samples into five different batches, similar to the observation made using the chromatographic profiles and HCA. The PCA further demonstrated that the simulated nyaope samples could still be discriminated even after 72 h of autosampler storage, confirming the stability of the samples once extracted into tertiary butyl alcohol.

## Conclusions

We describe, for the first time, a method for the comparative analysis of nyaope. Provided that the samples are correctly seized and stored[12], extracted into tertiary butyl alcohol[11], and analysed within 72 hours[11], this study demonstrates, for the first time, that quantitative comparisons of nyaope samples can be made. It also demonstrates, for the first time, that clustering techniques can be successfully applied to nyaope samples to identify different members of the same batch. This means that law enforcement agencies in the Southern African Development Community and beyond have, for the first time, the ability to analyse nyaope and compare forensic science data. This will allow distribution and trafficking routes to be identified and will assist in the determination of the origins of this drug. It does, however, require that decisions are made about how the samples will be treated prior to any investigative activity. It has been shown[12] that the extraction of the drugs for analysis should be made as soon as possible after samples are seized, and this requires planning before any police action.

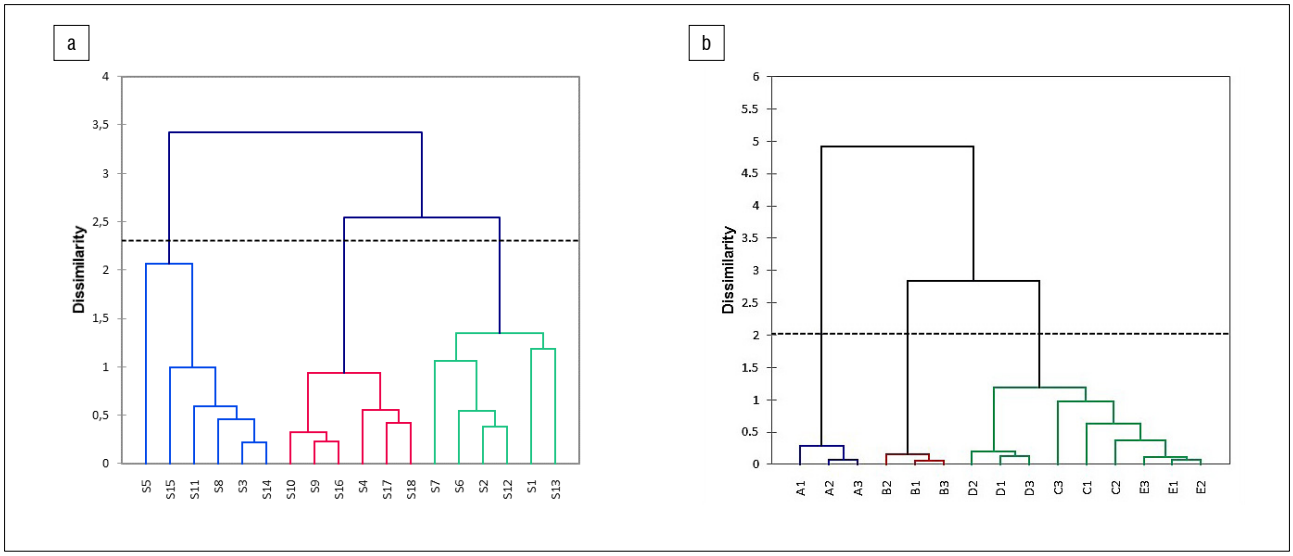



**Figure 3:** Dendrograms of (a) blind simulated nyaope samples and (b) casework samples analysed by hierarchical cluster analysis using unweighted linkage and Euclidean distance for the time interval t=72 h.

**Table 7:** Pearson correlation matrices for the simulated and casework samples

| Simulated samples | | | | | | | Case work | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Variables** | **Caffeine** | **Diamorphine** | **Efavirenz** | **Nevirapine** | **Phenacetin** | **$\Delta^9$-THC** | **Variables** | **Caffeine** | **Diamorphine** | **$\Delta^9$-THC** |
| **Caffeine** | 1 | 0.884 | 0.595 | -0.771 | 0.125 | 0.981 | **Caffeine** | 1 | 0.984 | -0.710 |
| **Diamorphine** | 0.884 | 1 | 0.290 | -0.471 | 0.556 | 0.858 | **Diamorphine** | 0.984 | 1 | -0.687 |
| **Efavirenz** | 0.595 | 0.290 | 1 | -0.645 | -0.437 | 0.512 | **$\Delta^9$-THC** | -0.710 | -0.687 | 1 |
| **Nevirapine** | -0.771 | -0.471 | -0.645 | 1 | 0.412 | -0.801 | | | | |
| **Phenacetin** | 0.125 | 0.556 | -0.437 | 0.412 | 1 | 0.089 | | | | |
| **$\Delta^9$-THC** | 0.981 | 0.858 | 0.512 | -0.801 | 0.089 | 1 | | | | |

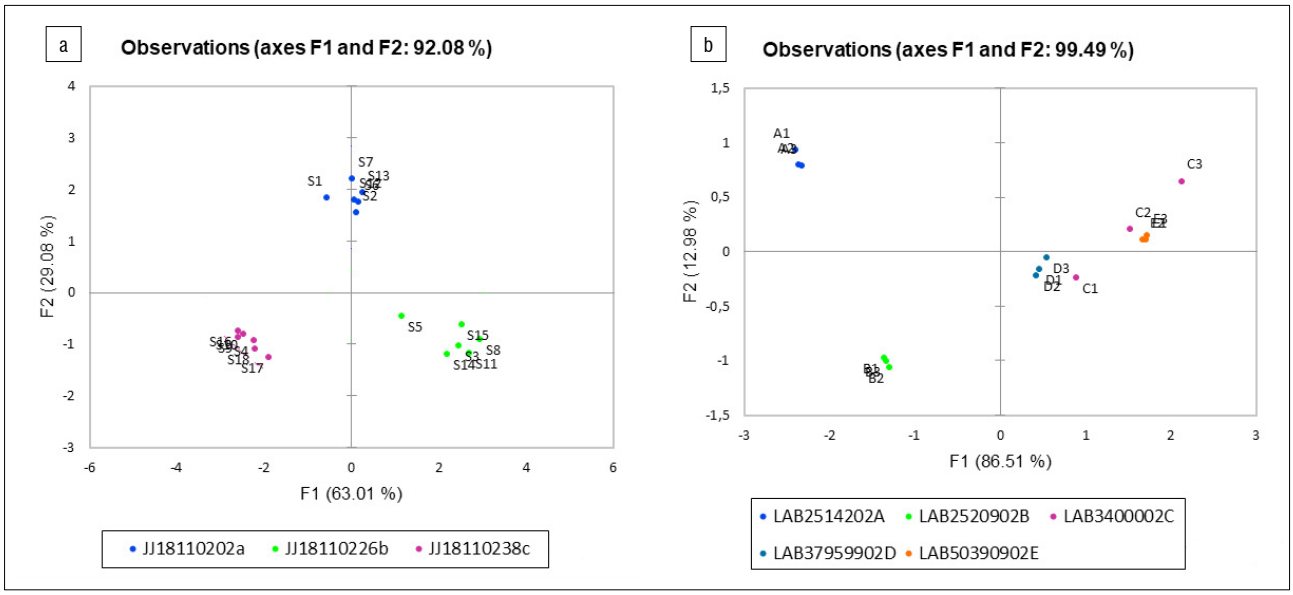



**Figure 4:** Observation axis for (a) blind simulated nyaope and (b) casework samples analysed using principal component analysis.

It is also interesting that whilst antiretrovirals have been reported to be present in nyaope, in the five casework samples analysed here, the antiretrovirals efavirenz and nevirapine were not identified. It may be that they were below the detection limit of the instrument or that they were indeed absent. However, the method does provide for the determination of these antiretrovirals where they are present at concentrations above the detection threshold. It should, however, be noted that other antiretrovirals cannot be detected by GC–MS, but that liquid chromatography – mass spectrometry (LC–MS) is a suitable method. However, LC–MS is not currently readily available to SAPS.

This method assists law enforcement and public health officials in a number of ways. It assists the law enforcement agencies in the identification and comparison of nyaope samples. It allows the establishment, for the first time, of a database on the composition of nyaope. It allows exchange of analytical data between jurisdictions, provided that the necessary quality control protocols are in place. It also facilitates the prosecution of trafficking offences. In terms of public health, it allows determination of the drugs present in nyaope and better public health information to be disseminated amongst the users of nyaope. In turn, they may choose, having this information, to avoid using this drug cocktail.

## Acknowledgements

We thank members of the European Network of Forensic Science Institutes (ENFSI) Drugs Working Group, and especially Joana Alegre, Portuguese Criminal Police, for helpful discussion and their interest in this work. We thank the University of the Free State, South Africa, for financial support and the South African Police Service (SAPS) for support in conducting this research and for allowing the findings to be disseminated.

## Competing interests

The research was supported by the South African Police Service Forensic Science Laboratory, of which P.M.M. is an employee. E.M.M. and M.D.C. have no conflicts of interest to declare.

## Authors' contributions

P.M.M.: Conceptualisation, methodology, data collection, writing – original draft, review and editing. E.M.M.: Supervision, writing – original draft, validation, review and editing. M.D.C.: Supervision, writing – original draft, validation, review and editing.

## References

1. Mokwena K. 'Consider our plight': A cry for help from nyaope users. Health SA Gesondheid. 2016;21:137–142. https://doi.org/10.1016/j.hsag.2015.09.001
2. Grelotti DJ, Closson EF, Smit JA, Mabude Z, Matthews LT, Safren SA, et al. Whoonga: Potential recreational use of HIV antiretroviral medication in South Africa. AIDS Behav. 2014;18:511–518. https://doi.org/10.1007/s10461-013-0575-0
3. Mokwena KE. The novel psychoactive substance 'nyaope' brings unique challenges to mental health services in South Africa. Int J Emerg Ment Health Hum Resil. 2015;17(1):251–252. https://doi.org/10.4172/1522-4821.1000152
4. Monyakane MM-EM. A rehabilitative South African criminal law response to nyaope, drug addiction: A recommendation for health oriented nyaope drug weaning. Res Pediatr Neonatol. 2018;3(1):206–214. http://dx.doi.org/10.31031/rpn.2018.03.000554
5. Masombuka J. Children's addiction to the drug 'nyaope' in Soshanguve Township: Parents' experiences and support needs [thesis]. Pretoria: University of South Africa; 2013. http://hdl.handle.net/10500/11903
6. Thomas R, Velaphi S. Abuse of antiretroviral drugs combined with addictive drugs by pregnant women is associated with adverse effects in infants and risk of resistance. S Afr J Child Health. 2014;8(2):78–79. https://doi.org/10.7196/sajch.734
7. Meel R, Essop MR. Striking increase in the incidence of infective endocarditis associated with recreational drug abuse in urban South Africa. S Afr Med J. 2018;108(7):585–589. https://doi.org/10.7196/samj.2018.v108i7.13007
8. Chambers D. 10 die after destroying their hearts by mainlining nyaope. TimesLive South Africa. 2018 July 02. Available from: https://www.timeslive.co.za/news/south-africa/2018-07-02-10-die-after-destroying-their-hearts-by-mainlining-nyaope/
9. Recovery Direct. Cape Town's Top Whoonga / Nyaope Rehab Centre [webpage on the Internet]. No date [cited 2019 Jun 06]. Available from: https://www.recoverydirect.co.za/drug/whoonga-addiction/
10. Haysom S. Hiding in plain sight: Heroin's stealthy takeover of South Africa. ENACT Heroin Policy Brief. 2019 April 11. Available from: https://enactafrica.org/research/policy-briefs/hiding-in-plain-sight-heroins-stealthy-takeover-of-south-africa

11. Mthembi PM, Mwenesongole EM, Cole MD. Chemical profiling of the street cocktail drug 'nyaope' in South Africa using GC–MS II: Stability studies of the cannabinoid, opiate and antiretroviral components during sample storage. Forensic Sci Int. 2019;300:187–192. https://doi.org/10.1016/j.forsciint.2019.04.040

12. Mthembi PM, Mwenesongole EM, Cole MD. Chemical profiling of the street cocktail drug 'nyaope' in South Africa using GC–MS I: Stability studies of components of 'nyaope' in organic solvents. Forensic Sci Int. 2018;292:115–124. https://doi.org/10.1016/j.forsciint.2018.08.001

13. ICH Harmonised Tripartite Guideline: Validation of analytical procedures: Text and methodology, Q2 (R1) [document on the Internet]. c2005 [cited 2019 Jul 07]. Available from: http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q2_R1/Step4/Q2_R1__Guideline.pdf

14. UNODC. Guidance for the validation of analytical methodology and calibration of equipment used for testing of illicit drugs in seized materials and biological specimens. Vienna: United Nations; 2009.

15. Bonfilio R, Cazedey ECL, De Arańjo MB, Salgado HRN. Analytical validation of quantitative high-performance liquid chromatographic methods in pharmaceutical analysis: A practical approach. Crit Rev Anal Chem. 2012;2:87–100. https://doi.org/10.1080/10408347.2012.630926

16. Skoog DA, West DM, Holler FJ, Crouch SR. Fundamentals of analytical chemistry. 8th ed. Toronto: Brooks/Cole – Thomson Learning Inc.; 2004.

17. Ahmad UK, Muniandy Y, Hassan MS. Physical analysis and chemical profiling of illicit herbal cannabis using multivariate analysis. Malaysian J Forensic Sci. 2005;5(1):26–34. http://forensics.org.my/mjofs/pdf/fssmVol.5No.1/Article%2005.pdf

18. United Nations Office on Drugs and Crime. Recommended methods for the identification and analysis of cannabis and cannabis products. Vienna: United Nations Publications; 2009.

19. Peters FT, Drummer O H, Musshoff F. Validation of analytical methods. Forensic Sci Int. 2007;165(2–3):216–224. https://doi.org/10.1016/j.forsciint.2006.05.021

20. González O, Blanco ME, Iriarte G, Bartoloméd L, Maguregui MI, Alonso RM. Bioanalytical chromatographic method validation according to current regulations, with a special focus on the non-well defined parameters limit of quantification, robustness and matrix effect. J Chrom A. 2014;1353:10–27. https://doi.org/10.1016/j.chroma.2014.03.077

21. Kadiana N, Raju KSR, Rashid M, Malika MY, Taneja I, Wahajuddin M. Comparative assessment of bioanalytical method validation guidelines for pharmaceutical industry. J Pharm Biomed Anal. 2016;126:83–97. https://doi.org/10.1016/j.jpba.2016.03.052

22. Tarcomnicul I, Van Nuijs AL, Simons W, Bervoets L, Blust R, Jorens PG, et al. Simultaneous determination of 15 top-prescribed pharmaceuticals and their metabolites in influent waste water by reversed-phase liquid chromatography coupled to tandem mass spectrometry. Talanta. 2011;83:795–803. https://doi.org/10.1016/j.talanta.2010.10.045

23. Karolak S, Nefau T, Bailly E, Solgadi A, Levia Y. Estimation of illicit drug consumption by waste water analysis in Paris area (France). Forensic Sci Int. 2010;200(1–3):153–160. https://doi.org/10.1016/j.forsciint.2010.04.007

24. United Nations Office on Drugs and Crime – Laboratory and Scientific Section. Methods for impurity profiling of heroin and cocaine: Manual for use by national drug testing laboratories. Vienna: United Nations Publications; 2005.

25. Hennig C. Dissolution point and isolation robustness: Robustness criteria for general cluster analysis methods. J Multivar Anal. 2008;99(6):1154–1176. https://doi.org/10.1016/j.jmva.2007.07.002

26. Shlens J. A tutorial on principal component analysis [document on the Internet]. c2005 [cited 2020 Mar 22]. Available from: http://www.cs.cmu,edu/∼elaw/papers/pca.pdf

27. Kaiser H. A second generation little jiffy. Psychometrika. 1970;35(4):401–415. https://doi.org/10.1007/bf02291817

28. Jolliffe IT, Cadima J. Principal component analysis: A review and recent developments. Philos Trans R Soc A. 2016;374, Art. #20150202. http://dx.doi.org/10.1098/rsta.2015.0202

**AUTHORS:**
Kunle I. Olatayo[1]
Paul T. Mativenga[1,2]
Annlizé L. Marnewick[1]

**AFFILIATIONS:**
[1]Postgraduate School of Engineering Management, University of Johannesburg, Johannesburg, South Africa
[2]Department of Mechanical, Aerospace and Civil Engineering, School of Engineering, The University of Manchester, Manchester, United Kingdom

**CORRESPONDENCE TO:**
Kunle Olatayo

**EMAIL:**
kunleo@uj.ac.za

**DATES:**
**Received:** 17 Sep. 2020
**Revised:** 16 Apr. 2021
**Accepted:** 04 July 2021
**Published:** 29 Nov. 2021

**HOW TO CITE:**
Olatayo KI, Mativenga PT, Marnewick AL. Life cycle assessment of single-use and reusable plastic bottles in the city of Johannesburg. S Afr J Sci. 2021;117(11/12), Art. #8908. https://doi.org/10.17159/sajs.2021/8908

**ARTICLE INCLUDES:**
☒ Peer review
☐ Supplementary material

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

**EDITOR:**
Priscilla Baker

**KEYWORDS:**
plastic, life cycle assessment, PET, single-use bottle, reusable bottle

**FUNDING:**
None

# Life cycle assessment of single-use and reusable plastic bottles in the city of Johannesburg

Polyethylene terephthalate (PET) bottles of water have experienced huge growth in demand and sales in South Africa. This expansion in use creates challenges as well as opportunities for managing the life cycle impact. The properties that make PET desirable for fluid-containing bottles have also made it highly resistant to environmental biodegradation. Reusable plastic bottles are now marketed as a solution to reduce the impact of single-use plastic bottles. We assessed the life cycle impact of single-use PET bottles and an alternative, reusable PET bottle based on consumption patterns in South Africa and the material flow and supply chain in the urban environment. This robust consideration of local conditions is important in evaluating the life cycle impact. In an examination of 13 impact categories, the reusable PET bottle had lower impact than the single-use bottle in all the impact categories examined. The mass of PET bottle material required to deliver the water needs at any given time is a dominant factor on the environmental burden. Extending the life of reusable bottles and designing lighter weight bottles would reduce their life cycle impact. Information obtained in evaluating alternatives to plastic water bottles can be valuable for providing a foundation assessment for policymakers and plastic bottle manufacturers to make informed choices and to focus on improvements in life cycle impact.

**Significance:**

- The significant impact of the production phase in the life cycle of both single-use and reusable PET bottles confirms the need to design a much more lightweight bottle to reduce the mass of materials used in production.
- Another key consideration was the long transportation distance covered during the production phase, and the negative impact of current vehicular emissions. Municipalities and waste collectors should consider the use of low-carbon transport.
- This study highlights the value of extending the life of plastic bottles, as well as recycling for material recovery, remanufacturing and repurposing these bottles within the City.
- The use of fewer, larger single-use bottles compared with a greater number of smaller single-use bottles is discussed.

## Introduction

The plastic packaging sector has the highest percentage of plastic consumption globally, contributing to about 40% of plastic usage in 2014.[1] However, the majority of the plastics produced in this sector is single-use plastic.[2] The most widely used plastic for packaging is polyethylene terephthalate (PET) due to the combination of valuable properties such as strength, chemical and thermal stability, easy processing, durability, and cost effectiveness.[3] These valuable qualities make it desirable for the production of packaging for water.[4] The plastic bottle is a major application of PET[3], with PET constituting 62% of manufactured bottles globally[5].

Interestingly, the valuable properties that make PET desirable for bottle production also make it highly resistant to environmental biodegradation. It has been suggested that PET can accumulate in landfills for years, leading to major environmental and health concerns.[4,6,7] It has been reported that the plastic bottle is the most common single-use plastic packaging, most of which becomes waste after initial use.[8] According to the estimation of the South African Plastic Recycling Organisation[9], about 90% of plastic bottle waste goes to landfills. Most water bottling companies are still reluctant to consider recycling plastics as the cost of new plastic resin may be lower than the associated cost of recycling.[10] In South Africa, there is rising concern about mismanaged plastic packaging waste, as plastic pollution is on the increase.[11]

Considering the numerous benefits and corresponding environmental burden of plastic PET bottles, there is an urgent need for a better understanding of their life cycle, fate and pathways for any nation or community.[12,13] It is imperative to develop a locally sustainable solution to the challenges of production, accumulation and impacts of plastic bottles in the environment. Most practical solutions recommended include substituting plastic with other materials, recycling, and reuse. Many studies have examined the environmental impacts of different materials as substitutes for plastic packaging and bottles, and others have evaluated the impacts of different end-of-life waste management scenarios. Table 1 compares the aims, assessments and main findings of previous studies.

Based on Table 1 and prior research, it is suggested that when considering global warming potential, it cannot be assumed that glass bottles are more environmentally friendly than PET bottles.[14] The other studies in Table 1 also show that environmental burden varies according to choice of end-of-life options. These findings highlight the need to do a detailed life cycle assessment (LCA) of different packaging options.

**Table 1:** Focus and findings of life cycle assessment studies on environmental impacts of plastic products

| Reference | Purpose | Functional unit | Findings |
|---|---|---|---|
| Kouloumpis et al.[14] | Impact of substituting PET bottles with glass bottles | Total community consumption of PET bottles | Substitution could lead to significant increases in global warming potential and climate change |
| Abejon et al.[38] | Life cycle impact assessment of plastic crates and cardboard boxes packaging for the distribution of fruits and vegetables | 10 000 050 units of both fruit and vegetable packages | Reusable plastic crates had lower impacts than single-use cardboard boxes |
| Horowitz et al.[7] | Impact assessment of different bottles for bottled water: ENSO, PLA, recycled PET, and regular PET | Typical package of 12 bottles | ENSO bottles had the lowest impacts, followed by recycled PET bottles |
| Wager and Hischier[27] | Impacts of recycling and incinerating plastic residues from waste electrical and electronic equipment (WEEE) | 1 tonne of plastic residues from the management of WEEE | Recycling had less impact than the existing options of disposal and incineration |
| Chen et al.[28] | Impact assessment of recycling, incineration and landfill in the management of plastic wastes | 1 tonne of waste plastics | Mechanical recycling had better environmental impact than incineration and landfill |

For South Africa, a review of the different studies on plastic revealed that there is no known available comprehensive information on the environmental life cycle of plastic bottles. A number of studies in the country have focused on microplastic pollution in water resources[15,16] and policies and sustainability in plastic use[17-19]. The studies on microplastics showed the accumulation and high concentration of plastics on shores and beaches due to an increase in the urbanisation levels in areas around them and this pollution poses environmental and health risks to the environment, biodiversity and coastal communities. Also, the failure to consider key stakeholders in the policy implementation process of the Plastic Bags Regulation, which prescribes a minimum thickness of 24 $\mu$m and maximum printing of 25% of the surface area of the bag and a fixed minimal charge of ZAR0.46 per plastic bag[18], has been identified as a barrier to eliminating plastic shopping bags from the environment.

Therefore, in addressing the knowledge gaps regarding the environmental life cycle of plastic packaging in South Africa, and plastic bottles for water in particular, a LCA was conducted to comprehensively assess the environmental impact of plastic bottles in one of the most populated and commercialised cities in Africa, namely Johannesburg. The City of Johannesburg is sited at an elevation of about 1740–to 1810 metres in the Highveld, a plateau area of South Africa.[20] It is not a coastal city, neither is it built on a river or harbour; however, there are some streams and artificial lakes and canals.

In this study, we assessed the environmental impacts associated with the life cycle of single-use PET bottles and reusable PET bottles. Our assessment considered the local supply chain and use patterns, the different phases of production, use, and disposal, and the current and future role of plastic collection and recycling as driven by an informal waste collection sector. Single-use PET bottles are described as the plastic bottles that are mostly discarded and that become waste after the initial consumption of their contents.[21] Reusable PET bottles are the bottles designed to be stronger and that are used, washed, and reused.

These two types of plastic bottles, shown in Figure 1, were considered for this study because of their substantial use in South Africa.[22] The single-use PET bottle is one of the most common single-use plastic packaging.[23] Bottled water has constantly experienced huge growth in demand and sales in South Africa in recent years as a result of the continual water crisis plaguing the nation due to drought and water diseases[24] and issues with wide-scale availability of purified water. The bottled water industry makes a major contribution to the economy of South Africa as it generated about ZAR3 550 million in sales and employed 1800 people in South Africa in 2011.[25] For the reusable PET bottles, their reusability and the refill water initiative at different shopping outlets and petrol stations in the country support the continued demand for the product. There is also potential for increased growth of the PET bottle industry with the increase in the number of entrepreneurs entering the market. Determining the environmental impacts of these PET bottles is thus timely and important for South Africa. Furthermore, to our knowledge, no study has assessed and compared the environmental impacts of reusable bottles with single-use bottles. Most importantly, similar studies conducted globally cannot be applied directly to South Africa, as system parameters contribute significantly in determining the environmental impacts of products and processes in an individual country.

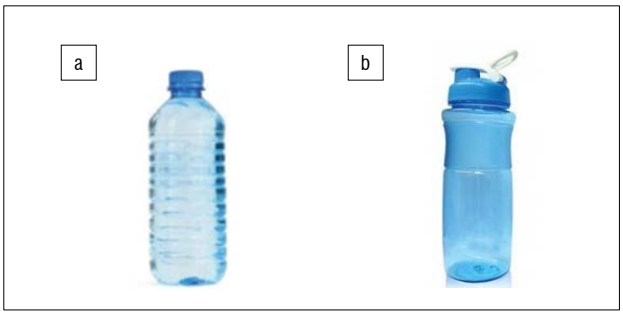



**Figure 1:** Samples of 0.5-L PET bottles: (a) single use and (b) reusable.

Life cycle assessment is an important method used to analyse and quantify the potential environmental impacts and resource consumption associated with a product system.[26] Evaluating alternative plastic water bottles provides a foundation assessment for policymakers and plastic bottle manufacturers to make informed choices and to focus on improvements in environmental performance.

## Methodology

The LCA method was used for this study. The four steps for conducting an LCA were followed, namely goal and scope definition, inventory analysis, impact assessment and interpretation.[26] Furthermore, sensitivity analysis was applied to evaluate the reliability of the results by varying some of the input data to determine the corresponding changes in the outputs. This helped to determine the significance and influences of selected data and evaluation methods on the LCA results.[27,28]

### Goal and scope definition

The purpose of our study was to comparatively analyse and quantify the resource consumption and environmental impacts associated with the use of single-use PET bottles and a probable environmentally friendly plastic bottle alternative, namely reusable PET bottles, in the City of Johannesburg, South Africa. Specifically, we evaluated the resource consumption and environmental impact associated with the different phases of production, use and disposal of single-use and reusable PET bottles. Sensitivity analysis was used to assess the impact of input data regarding the evaluation of resource use and environmental impacts. We then assessed (1) estimated impacts based on annual consumption of PET bottles and (2) implications of informal collection and recycling of PET bottle waste, which is a key factor in South Africa.

The system boundaries of the LCA define the processes within a product supply chain and account for time, space and functional unit.[29,30] The system boundaries of the plastic life cycle study included all activities and processes related to the production, use and disposal of plastic PET bottles in Johannesburg as illustrated in Figure 2. Life cycle activities such as raw material extraction and recycling were held constant as an assumption.[31,32]

The functional unit was defined as the delivery of bottled water to one person in Johannesburg for the period of one year. The functional unit was equivalent to a service life of a year for the reusable PET bottle, as informed by the findings of the survey by Tukur et al.[33] It has been suggested that the extended use of the reusable PET water bottle is not recommended due to the risk of discharge of antimony from PET bottles to the liquid contents.[34] For users, the deteriorating taste of water is also a factor. Considering that the South African market, with an estimated population of 58.8 million[35], consumed 617.3 million units of single-use PET bottled water in 2019[24], the functional unit of 10 single-use 0.5-L PET bottles was assumed for this study. This assumption is closely supported by a report of the South African National Bottled Water Association which estimated that 270 million litres of water are bottled annually.[36] It is noted that the average consumption in urban areas could be slightly higher than in other areas; however, there are limited data available for specific locations in the country. Similar assumptions have been made by other related studies.[14] Therefore, for this study, a single unit of a reusable 0.5-L PET bottle serves the same function as 10 units of single-use 0.5-L PET bottles.

### LCA inventory analysis

The data and information for the impact assessment were derived from actual measurements of bottles, study of use patterns, industry professionals and government agencies, and the review of a standard LCA database (EcoInvent 3.0) and scientific literature. The material inventory was based on the actual weight of bottles, water and packaging film and national statistics. The distances were estimated with the use of Google Maps, and electricity for production was supplied from the national grid. The scenarios for the manufacturing and transportation processes in this impact assessment were based on the actual PET bottle supply chain system investigated for the City of Johannesburg.

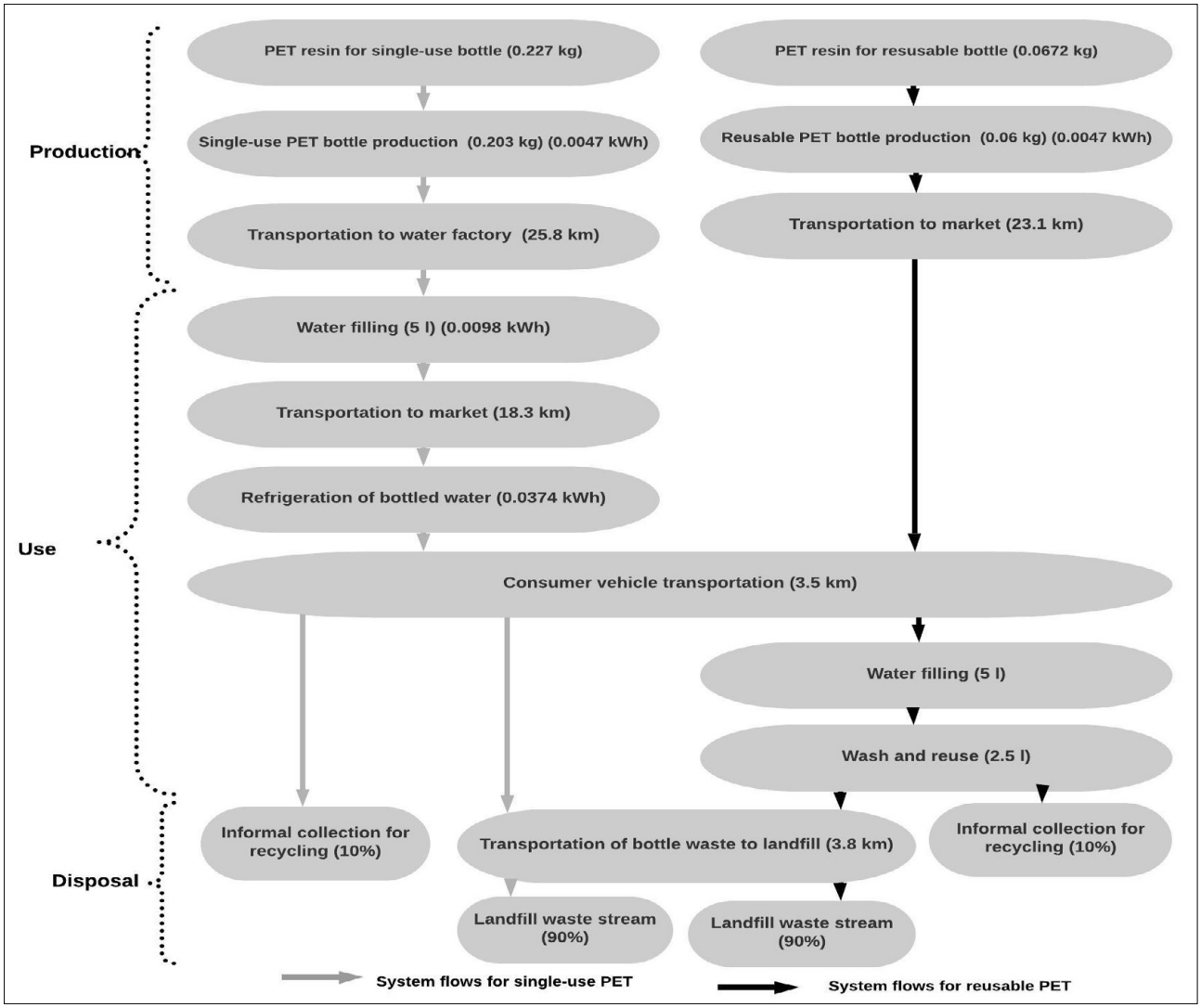



**Figure 2:** System boundaries of the life cycle of 10 single-use PET bottles and one reusable PET bottle as the functional units.

### Production phase of PET bottles

The PET bottles considered in this study were assumed to be produced from 100% virgin PET resins which are petroleum-based materials. Also, the bottle weight analysed included the weights of both the bottle cap (polypropylene) and label (low-density polyethylene). These assumptions are supported by previous research.[7] Actual measurements were conducted to derive the weights of the bottles. The weights of the single-use and reusable bottles are 0.0203 kg (bottle 0.018; cap 0.002; label 0.0003) and 0.060 kg (bottle 0.055; cap 0.005), respectively. The bottle production processes involve the plastic resin production, pre-forms production and blow moulding into bottles. Considering that a 1-kg PET bottle requires 1.12 kg PET resin[7] for production, a single-use bottle production consumes about 0.0227 kg of virgin PET resins. Similarly, a reusable bottle consumes about 0.0672 kg of virgin PET resins. The data for the blow-moulding process were extracted from the EcoInvent 3.0 database on SimaPro. As reported before, the study adopted a functional unit of per person annually for the bottled water needs of an individual in South Africa. This was 10 single-use 0.5-L PET bottles as supported by use patterns, and an equivalent functional unit of 1 reusable 0.5-L PET bottle per person annually.

Transportation of the bottle from the production plant was included in this phase. The transportation is a function of the weight of the goods and distance covered, calculated in kilogram-kilometre (kg-km) for all the routes.[14] The bottle manufacturing firm is located in Modderfontein, South Africa. The impact data for the truck transportation were retrieved from the EcoInvent 3.0 database. A distance of 25.8 km was estimated from the bottle factory in Modderfontein to the water factory in Randburg, South Africa, for water filling of single-use bottles. Also, transportation of reusable bottles from Modderfontein to market in Johannesburg's central business district was estimated at 23.1 km.

### Use phase of PET bottles

The impacts associated with the service life of the single-use bottle include the injection of water into the bottle (filling), transportation of bottled water to the market, refrigeration at the market, transportation to homes and initial consumption by end-users. The impact data for the water injection and refrigeration processes were modelled with the EcoInvent 3.0 database. The weight of water to fill the 0.5-L PET bottle is 0.516 kg. The plastic film to package the set of 10 bottles of water weighs 0.0184 kg. Both the weights of the water and plastic film were derived by actual measurement. Furthermore, the refrigeration of the bottled water was set at 0.0374 kWh[7], and the electricity for water filling at 0.0098 kWh (1 BTU = 0.000293071 kWh of electricity). An estimated distance of 18.3 km was covered to transport bottled water from the water factory in Randburg to the market in Johannesburg's central business district. For the reusable bottles, based on the actual supply chain, the impacts considered include transportation from the market to homes, initial consumption by filling with water, and the subsequent cleaning and reuse by end-users. Impact activities such as electricity for water filling and refrigeration were not considered as the real supply chain for reusable bottle in South Africa involves buying an empty bottle at the store and filling with water (which is usually tap or non-refrigerated water) at home or at designated places. The successive cleaning of the reusable bottle involves washing manually with a small volume of water, thus no electricity is involved. The water consumption for successive cleaning of a bottle is taken as half the bottle volume and the total is 2.5 L for the year. This assumption is supported by other studies.[37,38] Also, the consumer vehicle transportation distance from the market in Johannesburg's central business district to homes was estimated at 3.5 km.

### Disposal phase of PET bottles

In modelling the disposal phase of the two PET bottles, the single-use PET bottles are considered to be disposed of after the initial consumption of the contents[21], whilst the reusable PET bottles are assumed to be used repeatedly for one year and then disposed of, based on reports of random users. Water quality, bottle damage, aesthetics, new designs, etc. are expected to influence use patterns and shelf life. Approximately 90% of plastic bottle waste in South Africa ends up in landfills.[9] The average distance from the waste disposal location to the Robinson Landfill Site in Turffontein Stafford, South Africa is 3.8 km. Furthermore, plastic waste is mostly polymeric and does not degrade for hundreds of years.[32,39] However, plastic waste in landfill contributes to pollution and this was modelled in the software.

The LCA inventory of all the activities and processes involved in the different phases of the life cycles of the single-use and reusable PET bottles is further presented in Table 2. The total distances covered by the single-use and reusable PET bottles from production to disposal are 51.4 km and 30.4 km, respectively.

## *LCA Impact Assessment*

The IMPACT 2002+ assessment method in SimaPro 9 was used for the potential impact assessment of the PET bottles. This method has been used by other similar LCA studies[7,38,40], as it gives a comprehensive assessment of the processes examined, and is also among the current and up-to-date LCA methodologies. The impact categories evaluated in this LCA study were carcinogens, non-carcinogens, global warming, ozone layer depletion, aquatic eutrophication, aquatic ecotoxicity, terrestrial ecotoxicity, respiratory organics, respiratory inorganics, terrestrial acidification/nutrification, aquatic acidification, non-renewable energy (primary), and land occupation.

The carcinogens and non-carcinogens (kg $C_2H_3Cl$ eq.) are related to the formation of chemical compounds that affect human health and the ecosystem. Global warming (kg $CO_2$ eq.) is related to climate change, which is of public concern, and the environmental impact is assessed using greenhouse gases consisting of carbon dioxide, methane, nitrous oxide and other less prevalent gases. The ozone layer depletion (kg CFC11 eq.) impact is associated with the depletion of the ozone layer by chemical substances, as the ozone shields humans and organic matter from the ultraviolet radiation of the sun. The exposure to phosphorous compounds in the environment can be linked to aquatic eutrophication (kg $PO_4$ P-Lim) and this negatively affects plants and organisms through oxygen deprivation. Respiratory organics (kg $C_2H_4$ eq.) and inorganics (kg $PM_{2.5}$ eq.) are environmental impact categories related to the formation of tropospheric ozone and are a threat to health and quality of life.

**Table 2:** Functional unit inventory of PET bottles

| Item | Unit | Single use | Reusable |
|---|---|---|---|
| Functional unit of one person one-year bottled water supply in plastic bottle equivalent | P | 10 | 1 |
| Plastic bottle weight per functional unit | kg | 0.203 | 0.06 |
| Water volume per functional unit | kg | 5.16 | 5.16 |
| Bottle production | kWh | 0.0047 | 0.0047 |
| Transportation | km | 51.4 | 30.4 |
| Current landfill waste stream per functional unit | kg | 0.1827 | 0.054 |

Terrestrial and aquatic acidification (kg $SO_2$ eq.) has to do with the release of chemicals such as sulfur dioxides into the environment, causing lower than normal pH, which affects the acidity of the ecosystem. Aquatic and terrestrial ecotoxicity refers to substances that are poisonous to organisms in the ecosystem when emitted. Non-renewable energy (MJ) is primary energy such as coal and petroleum which cannot be reused, within a particular period, after the initial use. Land occupation ($m^2$org.arable) is associated with land mass that has the capability of being ploughed for useful purposes such as agriculture.

## Results and discussions

Table 3 shows the total impact values of the single-use and reusable PET bottles in all the impact categories. As shown, the reusable PET bottle had lower impact in all impact assessment categories evaluated than did the single-use PET bottle.

### *Comparative assessment of the life cycle phases*

The equivalent phases of the life cycles of the two different bottles were analysed and compared. The detailed values of the results for the different life cycle phases of production, use and disposal for the two PET bottles are shown in Table 4.

In all the impact categories, the impacts from the production, use and disposal phases of the single-use bottle were higher than the equivalent phases for the reusable bottle. The comparison of the use phases of the two PET bottles produced more interesting results. The percentage difference (or ratio) between the higher impact values of the single-use bottle and the lower impact values of the reusable bottle were very substantial in all impact categories in the use phases, unlike the differences between the other equivalent phases of the two bottles.

**Table 3:** Total life cycle assessment impacts of PET bottles

| Impact category | Unit | Single use | Reusable |
|---|---|---|---|
| Carcinogens | kg $C_2H_3Cl$ eq. | 0.00798 | 0.00234 |
| Non-carcinogens | kg $C_2H_3Cl$ eq. | 0.00583 | 0.00168 |
| Respiratory inorganics | kg $PM_{2.5}$ eq. | 2.90E-5 | 8.17E-5 |
| Ozone layer depletion | kg CFC-11 eq. | 2.421E-6 | 7.172E-7 |
| Respiratory organics | kg $C_2H_4$ eq. | 0.000455 | 0.000134 |
| Aquatic ecotoxicity | kg TEG water | 47.693 | 13.887 |
| Terrestrial ecotoxicity | kg TEG soil | 2.4043 | 0.5211 |
| Terrestrial acidification/nutrification | kg $SO_2$ eq. | 0.00722 | 0.00204 |
| Land occupation | $m^2$org.arable | 0.00529 | 0.00144 |
| Aquatic acidification | kg $SO_2$ eq. | 0.00195 | 0.00056 |
| Aquatic eutrophication | kg $PO_4$ P-Lim | 5.13E-5 | 1.49E-5 |
| Global warming | kg $CO_2$ eq. | 0.5128 | 0.1462 |
| Non-renewable energy (primary) | MJ | 15.8639 | 4.6089 |

**Table 4:** Life cycle assessment impacts of PET bottles in the different phases

| Impact category | Unit | Single-use | | | Reusable | | |
|---|---|---|---|---|---|---|---|
| | | Production | Use | Disposal | Production | Use | Disposal |
| Carcinogens | kg $C_2H_3Cl$ eq. | 0.0078 | 0.0001 | 4.93E-5 | 0.0023 | 1.51E-5 | 1.46E-5 |
| Non-carcinogens | kg $C_2H_3Cl$ eq. | 0.00542 | 0.00035 | 6.4E-5 | 0.00160 | 5.26E-5 | 1.89E-5 |
| Respiratory inorganics | kg $PM_{2.5}$ eq. | 0.00026 | 2.4E-5 | 6.64E-6 | 7.68E-5 | 2.94E-6 | 1.96E-6 |
| Ozone layer depletion | kg CFC-11 eq. | 2.42E-6 | 2.885E-9 | 7.79E-10 | 7.16E-7 | 4.46E-10 | 2.3E-10 |
| Respiratory organics | kg $C_2H_4$ eq. | 0.00044 | 1.0E-5 | 3.72E-6 | 0.00013 | 1.37E-6 | 1.1E-6 |
| Aquatic ecotoxicity | kg TEG water | 45.50 | 1.90 | 0.33 | 13.50 | 0.27 | 0.0979 |
| Terrestrial ecotoxicity | kg TEG soil | 0.9950 | 1.335 | 0.0652 | 0.2930 | 0.208 | 0.0193 |
| Terrestrial acidification/nutrification | kg $SO_2$ eq. | 0.00648 | 0.00060 | 0.00014 | 0.00193 | 7.18E-5 | 4.16E-5 |
| Land occupation | $m^2$org.arable | 0.00374 | 0.00085 | 0.00069 | 0.00111 | 0.00013 | 0.00021 |
| Aquatic acidification | kg $SO_2$ eq. | 0.00180 | 0.00012 | 2.95E-5 | 0.00054 | 1.22E-5 | 8.74E-6 |
| Aquatic eutrophication | kg $PO_4$ P-Lim | 4.84E-5 | 1.8E-6 | 1.03E-6 | 1.43E-5 | 2.77E-7 | 3.04E-7 |
| Global warming | kg $CO_2$ eq. | 0.471 | 0.030 | 0.012 | 0.140 | 0.002 | 0.0035 |
| Non-renewable energy (primary) | MJ | 15.300 | 0.500 | 0.078 | 4.550 | 0.039 | 0.023 |

This is possibly due to the additional amount of transportation in the use phase of the single-use PET bottle, as the single-use bottle experienced an initial distance of 18.3 km in the use phase when bottled water was transported from the water factory to the market in Johannesburg's central business district and another distance of 3.5 km for the consumer vehicle to transport from the market to the home. The only distance covered in the use phase by the reusable bottle was the consumer vehicle transportation from the market to the home (3.5 km). Transportation by motor vehicle increases fossil fuel use, which subsequently increases the amounts of chemical pollutants emitted into the environment.[7] Thus, single-use bottles are expected to have more impact. The sale of a higher number of single-use bottles in a pack could be a consideration in reducing impact from transport, provided other wider factors are also considered.

For the single-use bottle, the production phase had the highest impact in the impact categories, followed by the use phase and then the disposal phase. Also, for the reusable bottle, the production phase had the highest impact, followed by the use phase, with the disposal phase having the least impact. Manufacturers of reusable bottles should focus on the design and on reducing the environmental impact of manufacture and extended producer responsibility for, particularly, end-of-life waste management. The highest impact values experienced in the production phases of the two PET bottles compared to the other phases is consistent with studies in other countries.[14,38] The production phase of the single-use bottle has more impact than that of the reusable bottle due to the total weight of the 10 bottles for the functional unit.

The disposal phase of the reusable bottle has more impact than the use phase. This shows the importance of developing other end-of-life scenarios higher on the waste hierarchy. Instead of sending the reusable bottles to landfill, the uptake of bottle collection for recycling for material recovery, remanufacturing or repurposing should be promoted.

### *Sensitivity analysis of mass ratio*

Sensitivity analysis was conducted for this study by identifying parameters that could have an effect on the environmental impact results produced by the single-use PET bottle and reusable PET bottle. This helped to determine the reliability of the impact results and the consequence of alternative modelling. Considering that the LCA results established that the reusable bottle is more environmentally friendly, this analysis was done to determine the possibility of producing a more sustainable single-use PET bottle if certain variables in the life cycle of the bottle are reviewed. Additionally, in order to reduce or equalise the environmental impacts of the single-use PET bottle with that of the reusable bottle, the parameter of mass ratio of the required bottles was varied around their base values. This approach has been recommended by other studies.[14] Realising the recommendations would require advanced technological development. The single-use/reusable mass ratio exercise involves holding the mass of the reusable bottle constant while varying the mass of the single-use bottle around its base value, and then computing the impacts to determine when the impact category values of the two bottles become equal. This could further improve the environmental performance of the single-use PET bottle in the impact categories. The mass ratio of the bottles can be associated with the production phase, which had the greatest impact in the life cycles of the PET bottles. When the mass of the required single-use PET bottles is varied, the bottles could be equalised with the reusable bottle in most of the impact categories, except terrestrial ecotoxicity at a single-use/reusable mass ratio of 0.92, as shown in Figure 3. Also, it could perform better in all the impact categories at a mass ratio of 0.45, as presented in Figure 4. In practice this means that, (1) subject to technological advances, single-use bottles have to be lighter by circa 45%, or (2) lighter larger single-use bottles have to be promoted provided their total mass is lower than the smaller single-use bottle alternatives. Research, development and extending the useful life of reusable water bottles will make life cycle impacts of existing single-use bottles even less favourable.

The results further established that the mass or quantity of products from the production phase can greatly influence the environmental impact. The reduction in the mass of the single-use PET bottle could be achieved by introducing a much more lightweight design, which will result in reduced weight of the bottles, and will subsequently have less impact.

It needs to be noted that this study assessed the scenario of the two PET bottles as 100% virgin material. Other studies have already shown that the more the fraction of recycled PET, the better the performances in the impact categories.[7,41]

## Environmental impacts and waste management of bottles

The implications of the LCA results of the single-use and reusable PET bottles on the informal collection and recycling of these bottles for the City of Johannesburg are covered in this section.

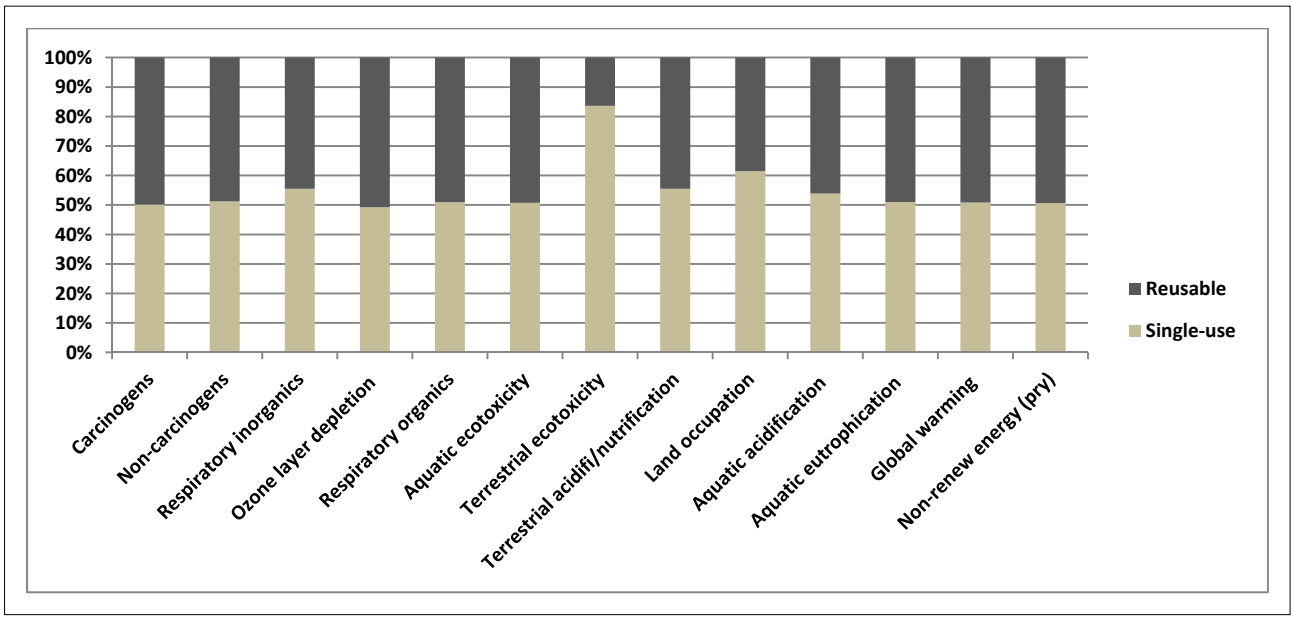



**Figure 3:** Equalisation of impacts for a single-use PET bottle at mass ratio of 0.92.

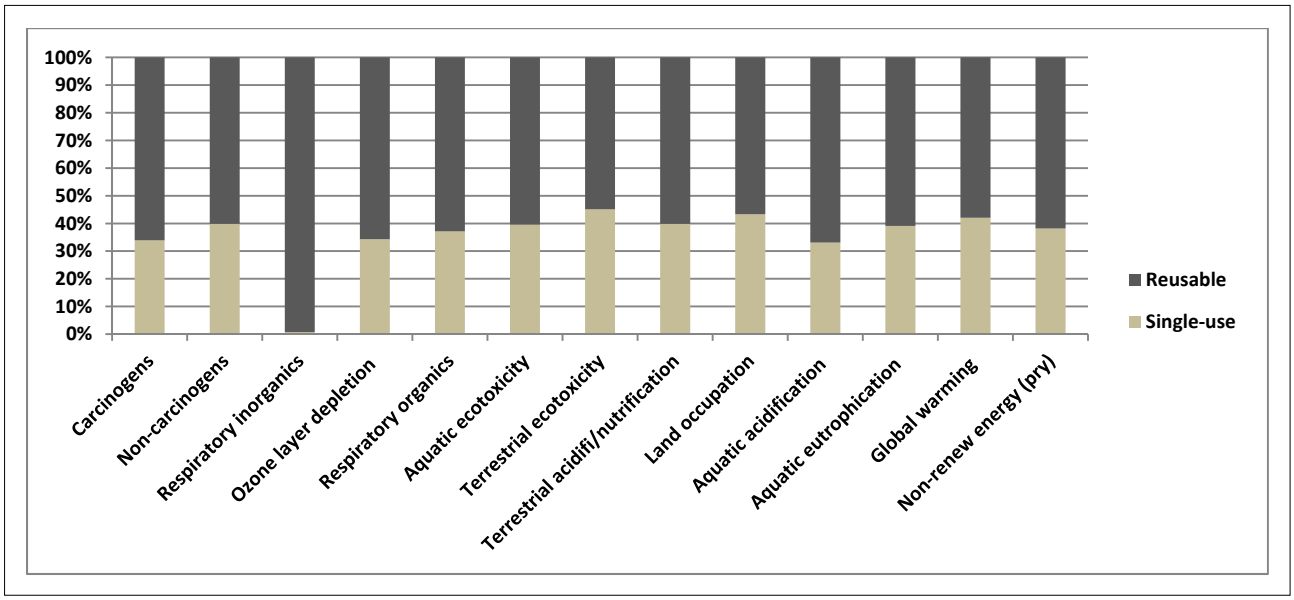



**Figure 4:** Better performance in impacts of a single-use PET bottle at mass ratio of 0.45.

**Table 5:** Environmental impacts avoided by the recycling rate scenarios

| Impact category | Unit | 10% Single use | Reusable | 46.3% Single use | Reusable |
|---|---|---|---|---|---|
| Carcinogens | kg $C_2H_3Cl$ eq. | 4.72E-5 | 1.40E-5 | 4.82E-5 | 1.43E-5 |
| Non-carcinogens | kg $C_2H_3Cl$ eq. | 6.09E-5 | 1.80E-5 | 6.23E-5 | 1.84E-5 |
| Respiratory inorganics | kg $PM_{2.5}$ eq. | 5.47E-6 | 1.62E-6 | 6E-6 | 1.78E-6 |
| Ozone layer depletion | kg CFC-11 eq. | 6.34E-10 | 1.88E-10 | 7E-10 | 2.07E-10 |
| Respiratory organics | kg $C_2H_4$ eq. | 2.24E-6 | 6.62E-7 | 2.91E-6 | 8.61E-7 |
| Aquatic ecotoxicity | kg TEG water | 0.297 | 0.088 | 0.313 | 0.093 |
| Terrestrial ecotoxicity | kg TEG soil | 0.0555 | 0.0164 | 0.0599 | 0.0177 |
| Terrestrial acidification/nutrification | kg $SO_2$ eq. | 0.000111 | 3.3E-5 | 0.000125 | 3.69E-5 |
| Land occupation | $m^2$org.arable | 0.000689 | 0.000204 | 0.000690 | 0.000204 |
| Aquatic acidification | kg $SO_2$ eq. | 2.49E-5 | 7.38E-6 | 2.70E-5 | 8E-6 |
| Aquatic eutrophication | kg $PO_4$ P-Lim | 9.59E-7 | 2.84E-7 | 9.89E-7 | 2.93E-7 |
| Global warming | kg $CO_2$ eq. | 0.0109 | 0.0032 | 0.0113 | 0.0033 |
| Non-renewable energy (primary) | MJ | 0.0654 | 0.0194 | 0.0709 | 0.0210 |

### *LCA impacts and informal collection and recycling of PET bottles*

The future increase in the consumption of plastic bottled water in the City of Johannesburg, with an estimated population of 5 635 000 as at 2019[42], will result in future increases in the volume of PET bottle waste generated. Regrettably, approximately 90% of plastic bottle waste in South Africa ends up in landfills, according to the analysis by the South African Plastic Recycling Organisation[9] applied in this study. However, several studies have established the preference for recycling over landfill, incineration, etc. in their assessments of the life cycle of different PET waste management options.[43,44] Interestingly, the recycling system of South Africa is incomplete without acknowledging the contributions of the informal sector.[45,46] Waste collectors from this sector are small-scale, self-employed agents, regularly found pushing their trolleys on the streets of Johannesburg and many urban areas of other developing countries.[47,48] For example, a total of 519 370 tonnes of plastic waste was collected for recycling in 2018 in South Africa.[49] The challenge is how to scale up the waste collection and to maximise the economic and social benefits for citizens.

In view of the importance of recycling and the contribution of the informal sector, we analysed two different recycling rate scenarios for the environmental impact that would be avoided by diverting the bottle waste from landfill to recycling in the disposal phase. These scenarios were the residual 10% of PET bottle recycled (assumed 90% landfill disposal), and the present 46.3% plastic recycling rate of South Africa obtained from Plastic SA[49]. It is crucial that the amount of 46.3% is examined given that not all plastic is collected. It is considered here as aspirational. The environmental impacts avoided through recycling are presented in Table 5.

The results show the significance of the environmental impact prevented by the two recycling rate scenarios. Also, the results support the inclusion of the informal sector in the official waste management system, as it contributes to environmental sustainability by reducing the volume of waste meant for landfill sites and by providing material for recycling.[45,50] In moving from recycling rates of 10% to 46.3%, the impacts avoided do not improve significantly. This suggests that recycling is not the major solution for reducing environmental impact, but that addressing the occurrence of plastic at source is the answer. However, it needs to be noted that, when considering the sustainability pillars, the informal recycling sector creates opportunities for revenue generation for society, provided the health and safety aspects of waste collection are also considered.

## LCA and sustainable design and manufacturing of PET bottles

General awareness about the potential impacts of the chemical components of PET bottles, and plastics in general, on human health and the environment is increasing. PET bottles are generally used as a container for liquids and the varied design features depend on the kind of liquid to be stored.[51] Innovations in the sustainable design and manufacture of PET bottles are being directed towards the reduction of the thickness or mass of plastic bottles.[51,52] The sustainable design of plastic or PET bottles is often defined in the area of circular economy or life cycle. The model of life cycle considers all the exposure, energy and emissions relating to the different phases of the life cycle of a product, including the extraction of raw material, production, use, and disposal and waste management. Taking into consideration the importance of life cycle in the design phase of plastic bottles will help manufacturers develop sustainable PET bottles that will positively affect the total environmental impacts at all life cycle phases of the bottle.

The sustainable design of PET bottles has its own challenges due to the complex nature of plastic, particularly the management of the end-of-life phase. As a result, DeCoster and Bateman[53] emphasised the importance and need for manufacturers to apply sustainable manufacturing approaches that will extend the lifetime of the product and reduce usage of resources, such as modularisation, design for closed loop, virtual manufacturing, product service system contract, upgradeable products and maintainable products.

## Conclusions

We comparatively analysed and quantified the environmental impacts associated with the life cycles of single-use PET bottles and reusable PET bottles in the City of Johannesburg in South Africa. Our study highlights the magnitude of the environmental problem facing the City caused by the huge consumption of PET bottles. There could also be a potential increase in this problem due to population growth and increased sales and demand for bottled water in the market. Our assessment established that the reusable PET bottle has a better environmental performance than the single-use PET bottle in the City of Johannesburg across all LCA impact categories. The primary reason for this difference is that more single-use bottles are needed for one year's supply of water compared to using one reusable bottle. The mass of material in the process of production contributes significantly to this greater impact. This also means that extending the life of the reusable bottle will make a positive contribution to reducing the life cycle impacts. Under the current conditions of usage of single-use bottles, use of lighter weight larger water bottles (rather than the alternative total number of smaller bottles) should be considered.

The significant impact of the production phase in the life cycle of the PET bottles confirms the need to design a much more lightweight single-use bottle and to reduce the mass of materials used in production. This was also established by the sensitivity analysis, which shows that the processes of manufacture and bottle forming associated with the production phase in the life cycle of a bottle product can be a major environmental cost, as the mass or quantity of material required can have a significant environmental impact. Also, more sustainable transportation of goods is required to lower the environmental impact.

Furthermore, we present the existing environmental impact in the City by analysing the total annual environmental impacts from the consumption of these PET bottles. Similarly, our analysis shows the importance and limitation of recycling and the beneficial role played by the informal sector in the waste management of plastic bottles by reducing the volume of plastic bottle waste meant for landfill sites and channelling it to recycling. However, recycling does not significantly reduce the life cycle impact of plastic bottle usage. A solution for addressing the demand for and consumption of plastic bottles at source is more desirable.

## Acknowledgements

We acknowledge SimaPro for the provision of LCA software for the study.

## Competing interests

We have no competing interests to declare.

## Authors' contributions

K.I.O.: Methodology, data analysis, software, writing – original draft preparation, investigation. P.T.M.: Conceptualisation, supervision, validation, writing – review and editing. A.L.M.: Conceptualisation, supervision, validation, writing – review and editing.

## References

1. Plastic Europe. Plastics: The facts 2014/15. An analysis of European latest plastics production, demand and waste data [document on the Internet]. c2015 [cited 2020 Jun 28]. Available from: https://www.plasticseurope.org/application/files/5515/1689/9220/2014plastics_the_facts_PubFeb2015.pdf
2. Muise I, Adams M, Cote R, Price GW. Attitudes to the recovery and recycling of agricultural plastics waste: A case study of Nova Scotia, Canada. Resour Conserv Recycl. 2016;109:137–145. https://doi.org/10.1016/j.resconrec.2016.02.011
3. Frigione M. Recycling of PET bottles as fine aggregate in concrete. Waste Manage. 2010;30:1101–1106. https://doi.org/10.1016/j.wasman.2010.01.030
4. Orset C, Barret N, Lemaire A. How consumers of plastic water bottles are responding to environmental policies. Waste Manage. 2017;61:13–27. https://doi.org/10.1016/j.wasman.2016.12.034
5. American Chemistry Council (ACC), Association of Plastic Recyclers (APR). 2018 US National post-consumer plastic bottle recycling report [document on the Internet]. c2019 [cited 2020 Apr 15]. Available from: https://www.plasticsmarkets.org/jsfcode/srvyfiles/w151_q2_sp88_0_2018_UNITED_STATES_NATIONAL_POSTCONSUMER_PLASTIC_BOTTLE_RECYCLING_REPORT.pdf
6. Revathi R, Kumar TR, Raman MS, Umanath B, Student UG. Reuse of bottles for wall construction and crafting. Int J Eng Sci. 2017;7(4):6693–6696.
7. Horowitz N, Frago J, Mu D. Life cycle assessment of bottled water: A case study of Green2O products. Waste Manage. 2018;76:734–743. https://doi.org/10.1016/j.wasman.2018.02.043
8. Heinrich Böll. Plastic atlas 2019: Facts and figures about the world of synthetic polymers [document on the Internet]. c2019 [cited 2020 May 21]. Available from: https://za.boell.org/en/2019/11/06/plastic-atlas-facts-and-figures-about-world-synthetic-polymers
9. South African Plastic Recycling Organisation (SAPRO). Why recycle? [webpage on the Internet]. c2019 [cited 2020 May 22]. Available from: https://www.plasticrecyclingsa.co.za/why-recycle/
10. Intagliata C. Does recycling plastic cost more than making it? [webpage on the Internet]. c2012 [cited 2020 May 17]. Available from: https://www.livescience.com/32231-does-recycling-plastic-cost-more-than-making-it.html
11. Ryan P, Perold V, Osborne A, Moloney C. Consistent patterns of debris on South African beaches indicate that industrial pellets and other mesoplastic items mostly derive from local sources. Environ Pollut. 2018;238:1008–1016. https://doi.org/10.1016/j.envpol.2018.02.017
12. Sonnemann G, Valdivia S. Medellin Declaration on marine litter in life cycle assessment and management. Int J Life Cycle Assess. 2017;22:1637–1639. https://doi.org/10.1007/s11367-017-1382-z

13. Boucher J, Faure F, Pompini O, Plummer Z, Wieser O, Felippe de Alencastro, L. (Micro) plastic fluxes and stocks in Lake Geneva basin. Trends Anal Chem. 2019;112:66–74. https://doi.org/10.1016/j.trac.2018.11.037

14. Kouloumpis V, Pell RS, Correa-Cano ME, Yan X. Potential trade-offs between eliminating plastics and mitigating climate change: An LCA perspective on polyethylene terephthalate (PET) bottles in Cornwall. Sci Total Environ. 2020;727(138681):1–10. https://doi.org/10.1016/j.scitotenv.2020.138681

15. Naidoo T, Glassom D, Smith AJ. Plastic pollution in five urban estuaries of KwaZulu-Natal, South Africa. Mar Pollut Bull. 2015;101(1):473–480. https://doi.org/10.1016/j.marpolbul.2015.09.044

16. Verster C, Minnaar K, Bouwman H. Marine and freshwater microplastic research in South Africa. Int Environ Assess Manage. 2017;13(3):533–535. https://doi.org/10.1002/ieam.1900

17. Nhamo G. Waste management policy implementation in South Africa: An emerging stakeholder participation paradox. South Afr J Environ Educ. 2003;20:39–52.

18. Dikgang J, Leiman A, Visser M. Analysis of the plastic bag levy in South Africa. Resour Conserv Recycl. 2012;66:59–65. https://doi.org/10.1016/j.resconrec.2012.06.009

19. Babayemi JO, Nnorom IC, Osibanjo O, Weber R. Ensuring sustainability in plastic use in Africa: Consumption, waste generation, and projections. Environ Sci Eur. 2019;31(60):1–20. https://doi.org/10.1186/s12302-019-0254-5

20. Campbell JT. Johannesburg South Africa [webpage on the Internet]. c2019 [cited 2020 Aug 04]. Available from: https://www.britannica.com/place/Johannesburg-South-Africa

21. Ma X, Park C, Moultrie J. Factors for eliminating plastic in packaging: The European FMCG experts' view. J Clean Prod. 2020;256(120492):1–20. https://doi.org/10.1016/j.jclepro.2020.120492

22. PET Plastic Recycling South Africa (PETCO). Plastic bottle recycled tonnage grown by 822% since 2005 [webpage on the Internet]. c2017 [cited 2020 May 22]. Available from: https://petco.co.za/plastic-bottle-recycled-tonnage-grown-822-since-2005/

23. Jambeck J, Geyer R, Wilcox C, Siegler T, Perryman M, Andrady A, et al. Plastic waste inputs from land into the ocean. Science. 2015;347(6223):768–771. https://doi.org/10.1126/science.1260352

24. Euromonitor International Limited. Bottled water in South Africa [webpage on the Internet]. c2020 [cited 2020 Jun 05]. Available from: https://www.euromonitor.com/bottled-water-in-south-africa/report

25. Metcalf C. Bottled water industry vital player in SA economy [document on the Internet]. c2012 [cited 2020 May 29]. Available from: https://www.environment.co.za/sustainable-green-business-news/bottled-water-industry-vital-player-in-sa-economy.html

26. ISO 14040. Environmental management – life cycle assessment principles and framework. European Standard EN ISO 14040 International Standardization Organization (ISO) [document on the Internet]. c2006 [cited 2020 Jun 22]. Available from: https://www.iso.org/standard/37456.html

27. Wager PA, Hischier R. Life cycle assessment of post-consumer plastics production from waste electrical and electronic equipment (WEEE) treatment residues in a Central European plastics recycling plant. Sci Total Environ. 2015;529:158–167. https://doi.org/10.1016/j.scitotenv.2015.05.043

28. Chen Y, Cui Z, Cui X, Liu W, Wang X, Li X, et al. Life cycle assessment of end-of-life treatments of waste plastics in China. Resour Conserv Recycl. 2019;146:348–357. https://doi.org/10.1016/j.resconrec.2019.03.011

29. Eriksson O, Frostell B, Bjorklund A, Assefa G, Sundqvist JO, Granath J, et al. ORWARE – A simulation tool for waste management. Resour Conserv Recycl. 2002;36(4):287–307. https://doi.org/10.1016/S0921-3449(02)00031-9

30. Antelava A, Damilos S, Hafeez S, Manos G, Al-Salem SM, Sharma BK, et al. Plastic solid waste (PSW) in the context of life cycle assessment (LCA) and sustainable management. Environ Manag. 2019;64:230–244. https://doi.org/10.1007/s00267-019-01178-3

31. Garfi M, Cadena E, Sanchez-Ramos D, Ferrer I. Life cycle assessment of drinking water: Comparing conventional water treatment, reverse osmosis and mineral water in glass and plastic bottles. J Clean Prod. 2016;137:997–1003. http://dx.doi.org/10.1016/j.jclepro.2016.07.218

32. Arena U, Mastellone ML, Perugini F. Life cycle assessment of a plastic packaging recycling system. Int J Life Cycle Assess. 2003;8(2):92–98. https://doi.org/10.1007/BF02978432

33. Tukur A, Sharp L, Stern B, Tizaoui C, Benkreira H. PET bottle use patterns and antimony migration into bottled water and soft drinks: The case of British and Nigerian bottles. J Environ Monit. 2012;14(4):1237–1247. https://doi.org/10.1039/C2EM10917D

34. Shotyk W, Krachler M, Chen B. Contamination of Canadian and European bottled waters with antimony from PET containers. J Environ Monit. 2006;8(2):288–292. https://doi.org/10.1039/b517844b

35. Stats SA. Mid-year population estimates 2019 [document on the Internet]. c2019 [cited 2020 Jul 11]. Available from: https://www.statssa.gov.za/publications/P0302/P03022019.pdf

36. South African National Bottled Water Association (SANBWA). About bottled water: Fact sheet [webpage on the Internet]. No date [cited 2020 May 15]. Available from: http://www.sanbwa.org.za/water_factsheet.asp

37. Landi D, Germani M, Marconi M. Analysing the environmental sustainability of glass bottles reuse in an Italian wine consortium. Proceedings of the 26th CIRP Life Cycle Engineering (LCE) Conference. Procedia CIRP. 2019;80:399–404. https://doi.org/10.1016/j.procir.2019.01.054

38. Abejon R, Bala A, Vazquez-Rowe I, Aldaco R, Fullana-i-Palmer P. When plastic packaging should be preferred: Life cycle analysis of packages for fruit and vegetable distribution in the Spanish peninsular market. Resour Conserv Recycl. 2020;155(104666):1–8. https://doi.org/10.1016/j.resconrec.2019.104666

39. Bez J, Heyde M, Goldhan G. Waste treatment in product specific life cycle inventories. Int J Life Cycle Assess. 1998;3(2):100–105. https://doi.org/10.1007/BF02978497

40. Jolliet O, Margni M, Charles R, Humbert S, Payet J, Rebitzer G, et al. IMPACT 2002+: A new life cycle impact assessment methodology. Int J Life Cycle Assess. 2003;8:324–330. https://doi.org/10.1007/BF02978505

41. Zhang R, Ma X, Shen X, Zhai Y, Zhang T, Ji C, et al. PET bottles recycling in China: An LCA coupled with LCC case study of blanket production made of waste PET bottles. J Environ Manag. 2020;260(110062):1–10. https://doi.org/10.1016/j.jenvman.2019.110062

42. Macrotrends LLC. South Africa metro area population 1950-2020 [webpage on the Internet]. c2020 [cited 2020 Jun 07]. Available from: https://www.macrotrends.net/cities/22486/johannesburg/population

43. Grant T, James KL, Lundie S, Sonneveld K. Stage 2 report for life cycle assessment for paper and packaging waste management scenarios in Victoria [document on the Internet]. c2001 [cited 2020 Jul 28]. Available from: https://docplayer.net/144697950-Stage-2-report-for-life-cycle-assessment-for-paper-and-packaging-waste-management-scenarios-in-victoria.html

44. Chilton T, Burnley S, Nesaratnam S. A life cycle assessment of the closed-loop recycling and thermal recovery of post-consumer PET. Resour Conserv Recycl. 2010;54(12):1241–1249. https://doi.org/10.1016/j.resconrec.2010.04.002

45. Langenhoven B, Dyssel M. The recycling industry and subsistence waste collectors: A case study of Mitchells Plain. Urban Forum. 2007;18(1):114–132. https://doi.org/10.1007/BF02681233

46. Schenck CJ, Blaauw PF. Living on what others throw away: A preliminary exploration into the socio-economic circumstances of people collecting and selling recyclable waste. UNISA's The Bright Site Project [document on the Internet]. c2010 [cited 2020 Jul 12]. Available from: https://www.researchgate.net/publication/269100683_Living_on_what_others_throw_away_An_exploration_of_the_socio_economic_circumstances_of_people_collecting_and_selling_recyclable_waste

47. Hayami Y, Dikshit AK, Mishra SN. Waste pickers and collectors in Delhi: Poverty and environment in an urban informal sector. J Develop Stud. 2007;42(1):41–69. https://doi.org/10.1080/00220380500356662

48. Schenck R, Blaauw PF. The work and lives of street waste pickers in Pretoria – A case study of recycling in South Africa's urban informal economy. Urban Forum. 2011;22:411–430. https://doi.org/10.1007/s12132-011-9125-x

49. Plastic SA. Key statistics from latest recycling report [webpage on the Internet]. c2019 [cited 2020 May 23]. Available from: https://www.plasticsinfo.co.za/2019/08/31/key-statistics-from-latest-recycling-report/

50. Samson M. Reclaiming reusable and recyclable materials in Africa: A critical review of English language literature. WIEGO working paper (Urban Policies) No. 16 [document on the Internet]. c2010 [cited 2020 May 14]. Available from: https://www.wiego.org/sites/default/files/publications/files/Samson_WIEGO_WP16.pdf

51. Steenis ND, Van der Lans IA, Herpen E, Van Trijp HCM. Effects of sustainable design strategies on consumer preferences for redesigned packaging. J Clean Prod. 2018;205:854–865. https://doi.org/10.1016/j.jclepro.2018.09.137

52. Mcharek M, Hammadi M, Azib T, Laroucib C, Choley JY. Collaborative design process and product knowledge methodology for mechatronic systems. Comput Ind. 2019;105:213–228. https://doi.org/10.1016/j.compind.2018.12.008

53. De Coster R, Bateman R. Sustainable product development strategies: Business planning and performance implications. Proc IMechE B J Eng Manuf. 2012;226(10):1665–1674. https://doi.org/10.1177/0954405412455123

**AUTHOR:**
Abdullahi A. Yusuf[1]

**AFFILIATION:**
[1]Social Insects Research Group, Department of Zoology and Entomology, University of Pretoria, Pretoria, South Africa

**CORRESPONDENCE TO:**
Abdullahi Yusuf

**EMAIL:**
abdullahi.yusuf@up.ac.za

**DATES:**
**Received:** 16 Dec. 2020
**Revised:** 24 June 2021
**Accepted:** 05 July 2021
**Published:** 29 Nov. 2021

**HOW TO CITE:**
Yusuf AA. Determination of alcohols in hand sanitisers: Are off-the-shelf hand sanitisers what they claim to be? S Afr J Sci. 2021;117(11/12), Art. #9328. https://doi.org/10.17159/sajs.2021/9328

**ARTICLE INCLUDES:**
☒ Peer review
☐ Supplementary material

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

**EDITOR:**
Priscilla Baker

**KEYWORDS:**
alcohol contents, ABHRs, COVID-19, SARS-CoV-2, homemade hand sanitisers, gas chromatography

**FUNDING:**
South African National Research Foundation, Alexander von Humboldt Foundation

# Determination of alcohols in hand sanitisers: Are off-the-shelf hand sanitisers what they claim to be?

Transmission of SARS-CoV-2 – the causative agent of COVID-19 – can be prevented through non-pharmaceutical interventions such as observing proper hand hygiene using alcohol-based hand rubs/sanitisers (ABHRs) as recommended by the WHO and local health authorities. However, this recommendation has led to high demand for ABHRs and proliferation of sub-standard products, which do not contain the recommended amount of alcohol. Fifty products of different origins and formulations obtained off-the-shelf and in public places in and around Pretoria (South Africa) were analysed for their alcohol content using gas chromatography. Ethanol was the most common alcohol used in the products, followed by isopropanol. Only 21 (42%) of the products analysed contained at least 70% alcohol; of these only 14 (28%) met the WHO recommended 80±5% alcohol content to have a virucidal effect on SARS-CoV-2. Of the 41 commercial off-the-shelf products analysed, 27 (66%) contained less than 70% alcohol in comparison to 13% of homemade products. Only 18% of gel products contained 70% alcohol, compared with 47% for liquid-based products. Most of the products did not contain the appropriate or correct declaration as recommended by the South African National Standards (SANS 289 and 490). The proliferation of substandard ABHRs is of great public health concern and calls for stricter regulations and enforcement in order to protect consumers, their rights and well-being during and post the COVID-19 pandemic period. However, in the interim, formulation of ABHRs using the WHO guidelines should be mandatory, as such formulations, when made correctly, do have the required virucidal effect against SARS-CoV-2.

**Significance:**

- Commercial, off-the-shelf and public hand sanitisers were analysed to determine whether they contained enough alcohol to be efficacious virucides as recommended by the WHO.
- The majority of the products analysed were substandard, did not contain the recommended amount of alcohols and were not labelled correctly according to local and international standards.
- Homemade products conformed to a greater degree to the WHO standards for alcohol-based hand sanitisers. It is evident from these results that there is a need to monitor the manufacture of off-the-shelf products to ensure compliance and to assure consumers that products offer the required protection against SARS-CoV-2.

## Introduction

In December 2019, an infectious disease named COVID-19 caused by a novel coronavirus (SARS-CoV-2)[1] was first identified in Wuhan, Hubei Province, China[2]. By the beginning of 2020, COVID-19 had rapidly spread around the world, leading to it being declared a pandemic by the World Health Organization (WHO) on 11 March 2020. By December 2020, a year later, global COVID-19 cases had reached 65 million, with 1.5 million deaths, of which 800 000 and 21 000 cases and deaths were from South Africa, respectively[2] – making South Africa among the countries with the highest per capita number of COVID-19 cases. Interrupting the chain of virus transmission using non-pharmaceutical interventions such as observing physical distancing, wearing of a face mask and maintaining good hand hygiene (washing hands with soap or the use of alcohol-based hand rubs/sanitisers (ABHR)) as recommended by the WHO and national health agencies remain the primary prevention options, especially given the slow pace of vaccination and the emergence of genetic variants of SARS-CoV-2. The recommendation that good hand hygiene should be practised led to stockpiling and hoarding of emergency supplies of hand sanitisers around the world – a phenomenon termed 'pandemic pantries' that resulted in the disappearance of these products from supermarket shelves.[3] After the rush, and with easing of lockdown restrictions, the return to work, and legislation that made the provision of hand sanitisers at public places mandatory, many manufacturers, including chemical industries, breweries and perfumeries, began producing ABHRs.[4] This demand drove the global hand sanitiser market valued at USD2.7 billion in 2019 up to USD3.3 billion in 2020, which is projected to reach USD13.7 billion by 2027.[5] In South Africa, production facilities for the raw materials needed to make hand sanitisers experienced a nearly 400% increase in demand, opening the door to the introduction of substandard products, exploitation (inflated prices) and corruption leading to estimated overpricing to the level of ZAR66 million.[6]

The effectiveness of an ABHR depends on the type of alcohol it contains, the concentration and quantity applied to hands as well as the duration of exposure.[7] ABHRs typically contain isopropyl alcohol (isopropanol), ethyl alcohol (ethanol), n-propanol or a combination of these alcohols.[8] However, the two formulations recommended by the WHO are: Formulation I containing 80±5 % ethanol (v/v) and Formulation II containing 75±5 % isopropanol.[9] Thus far, only ABHR formulations containing alcohols in the recommended concentrations by the WHO are shown to be effective against enveloped viruses including SARS-CoV-2.[10-12] As effectiveness of an ABHR depends on its alcohol content, quality control is essential in order to maintain the integrity of the product and ensure that consumers

are paying for and using products that have virucidal activity against COVID-19. It is therefore important to have easy and rapid methods to detect alcohols in ABHRs as well as to determine the alcohol content in commercial off-the-shelf ABHR products to ensure that the consumer is receiving an effective product. Hence, this study was undertaken with the following questions. Do the ABHRs sold or formulated in South Africa and those available in public places contain the required alcohol content to qualify as virucides? Are there differences between formulations (liquids and gels) in their alcohol content? Lastly, are ABHRs labelled appropriately as required by WHO and local standards set out in the South African National Standards SANS 289:2016 guide?[13] To answer these questions, a rapid and reproducible gas chromatographic method was used to determine the alcohol content of both off-the-shelf and homemade liquid- and gel-based ABHRs found and used in different public places in Pretoria. The findings here are discussed in the context of COVID-19 prevention and beyond.

## Materials and methods

### *Alcohol-based hand sanitisers*

Commercially available liquid- and gel-based ABHRs hereafter 'off the shelf' were purchased from stores or sampled from hand sanitising points at public places (schools, offices, shopping malls, restaurants, places of worship) in and around Pretoria, Gauteng Province, South Africa. Another set of ABHRs that were made based on WHO recommendations for hand rubs using locally available ingredients at home by persons or in the laboratory for the daily use of personnel (hereafter referred to as 'homemade') were solicited from and provided by volunteers. The ABHRs were sampled directly from the container (in the case of those purchased) or collected in 1.5-mL sterile Eppendorf tubes. Overall, 50 ABHR products were analysed, of which 38 were liquids, 11 were gels and 1 was a spray-based formulation.

### *Gas chromatographic analysis of alcohols in ABHRs*

For the analysis, a 6890 Agilent gas chromatograph fitted with a flame ionisation detector and an Agilent HP-INNOWax polyethylene glycol (model number Agilent 19091N-102) column (25 m × 200 $\mu$m × 0.20 $\mu$m) was used. Samples were injected in the split mode with a split ratio of 30:1 at 140 °C and a pressure of 24.14 psi. The temperature of the flame ionisation detector was set at 200 °C, the flow rates for hydrogen, air and nitrogen were set at 40, 150 and 25 mL/min, respectively, and that of the carrier gas helium was set at 1.4 mL/min. The oven was programmed as follows: 35 °C for 1 min, increased by 5 ° per min to 40 °C, held for 1 min ramped at 15 °C to 75 °C with a 1 min hold and finally increased at 25 °C per min to 220 °C. To validate the method, a mixture of seven alcohols containing methanol, ethanol, isopropanol, pentanol, isobutanol, butanol and octanol was analysed. Thereafter, 1 $\mu$L of each ABHR was analysed on the gas chromatograph and the alcohol content quantified using an external calibration curve made up of either ethanol or isopropanol (the two main alcohols recommended by the WHO for use in ABHRs) in the range of 10–100% (v/v). The concentration of either ethanol or isopropanol was determined using the equation of the trend line , The equation of the trend line $y=mx+c$, where y = measured peak intensity of alcohol, m = gradient, x = unknown concentration of alcohol and c = intercept. As a reference and quality control standard check, a 91% ethanol standard was analysed with each batch of samples. All samples were analysed in duplicate and the composition of alcohols presented as mean proportions ± standard errors. Comparisons between ABHR formulation types and sources were made using a chi-square ($X^2$) test of proportions using the Proc FREQ command in the statistical software SAS version 9.4 (SAS Inc, USA).

## Results

### *Direct method for analysing alcohols in ABHRs*

Using a direct gas chromatographic method for the analysis of alcohols in ABHRs, a mixture of seven alcohols was separated in less than 11 min (Figure 1). The two main alcohols used in ABHRs, ethanol and isopropanol, eluted from the column at 2.430 min and 2.530 min, respectively (Figure 2a and 2b).

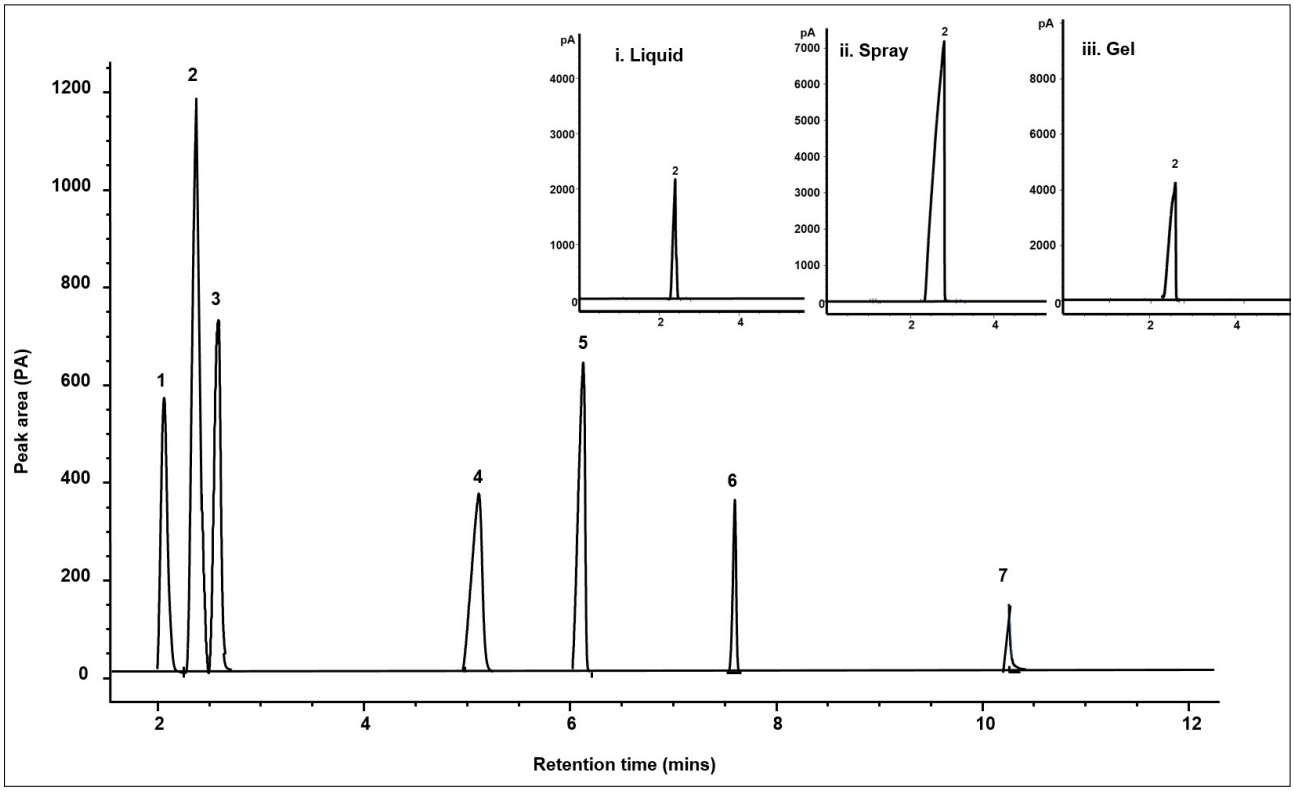



**Figure 1:** Separation of a mixture of seven alcohols – methanol (1), ethanol (2), isopropanol (3), pentanol (4), isobutanol (5), butanol (6) and octanol (7) – in less than 11 min on a 25-m Agilent HP-INNOwax polyethylene glycol capillary column. Inserts: Representative chromatograms of (i) liquid-, (ii) spray- and (iii) gel-based alcohol-based hand sanitisers.

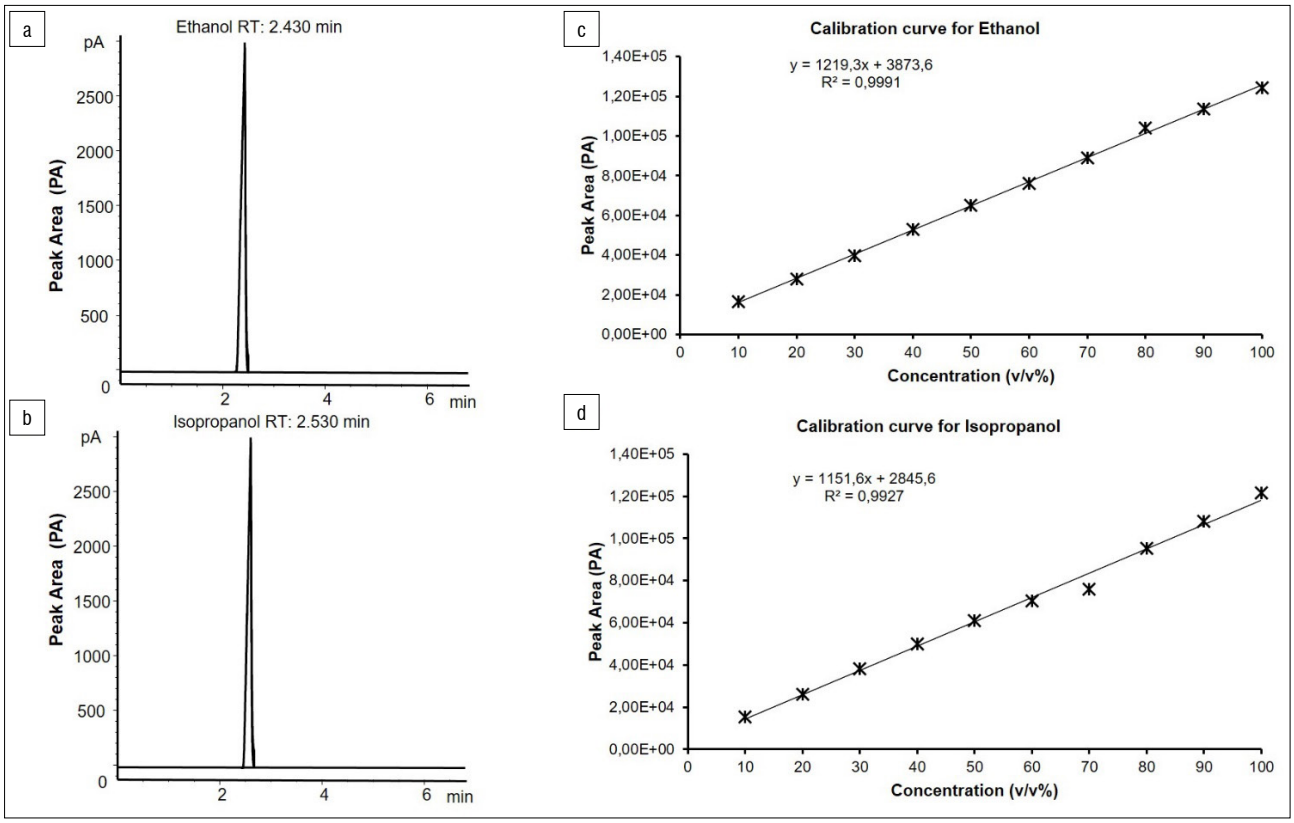



**Figure 2:** Elution times for ethanol (a) and isopropanol (b) and their respective calibration curves (c and d) indicating the accuracy of detection ($R^2$) and the regression parameters of the line $y=mx+c$.

Calibration curves prepared for both alcohols containing 10–100% (v/v) show a strong fit with coefficient of determination $R^2$-values of 0.999 and 0.992 for ethanol and isopropanol, respectively (Figure 2c and 2d). The 91% ethanol reference standard consistently had peak intensities with a mean of 114 716±190 over ten runs.

### *Alcohols and their composition in ABHRs*

Ethanol was the main alcohol found in the ABHRs with the exception of products 37 and 43 which contained isopropanol, and product 49 with a combination of isobutanol and ethanol (Table 1). Methanol was not detected in any of the ABHRs. Only 14 (28%) of the 50 ABHRs analysed contained ≥75% alcohol (Figure 3a, Table 1). Looking at alcohol composition by formulation type, only 13 (34%) of the 28 liquids and the only spray-based formulation (product 12) contained ≥75% alcohol (Figure 3a, Table 1). Comparing the alcohol composition of ABHRs based on their sources (homemade vs off the shelf), 63% of homemade ABHRs met the ≥75% alcohol content criterion whilst only 21% of the off-the-shelf ABHRs met this criterion ($X^2=5.798$, d.f.=1, $p=0.0160$, Figure 3a).

When the minimum alcohol composition was lowered to 70%, only 21 (42%) of the 50 ABHRs met this criterion, of which 18 (47%) and 2 (18%) are liquid and gel formulations, respectively ($X^2=7.268$, d.f.=1, $p=0.0070$, Figure 3b). Only 14 (34%) of the off-the-shelf ABHRs contained at least 70% alcohol and 7 of the 8 homemade ABHRs (88%) met this criterion ($X^2=7.810$, d.f.= 1, $p=0.0052$, Figure 3b).

### *Declaration and appropriate labelling*

All the ABHRs analysed, with the exception of products 24, 29 and 30, had labels on which the type of alcohol(s) they contained were declared (Table 1). However, only 29 (58%) of the 50 products gave an indication of the composition of alcohols they contained (Table 1). Of these, 16 products (55%) did not contain the amount of alcohol as declared on the labels (Table 1). One product (product 37) which had isopropanol as its main component, contained up to 99%, which is above the recommended 70±5% for isopropanol-based ABHRs.

## Discussion

A reproducible and direct method capable of detecting and separating the most commonly used alcohols in less than 11 min was used for the analysis of alcohols in hand sanitisers. This gas chromatographic method reduces analysis time and the need for complex sample preparation and offers the potential to increase throughput. Gas chromatography has previously been used for the determination of alcohols in ABHRs and application notes and methods such as those by Dhandapani[14] and Berardi et al.[15] are available. However, in these methods, alcohols, especially ethanol and isopropanol, were detected only after 4 minutes and were not directly sampled, hence increasing analysis time and the potential of introducing contaminants during sample preparation. Other analytical methods used for the determination of alcohols in ABHRs include Fourier transformed infrared spectroscopy for gel-based hand sanitisers.[16]

Most of the products analysed contained ethanol as their active ingredient, with only two found to contain isopropanol and one product contained a mixture of ethanol and isobutanol. Ethanol was the preferred alcohol over propanols for ABHR formulations because of its superior virucidal activity and its skin tolerance.[17,18]

A vast majority of the products analysed did not contain alcohols in the compositions recommended by the WHO (80±5 v/v% for ethanol and 70±5 v/v% for isopropanol).[9] Most also did not contain the required 70% ethanol recommended by the US Centers for Disease Control and Prevention (CDC). Because alcohol content and concentrations are imperative[3,7] for a sanitiser to have virucidal activity, these findings suggest that there is widespread lack of adherence to the required composition.

**Table 1:** Alcohol-based hand sanitiser products, their sources, formulation, use, type of alcohol, declared composition, composition ($n=2$), and average composition relative to WHO standards found using gas chromatography

| Product | Source | Type of formulation | Use | Type of alcohol | Alcohol content as stated on label (%) | % Alcohol found ($n=2$) | More or less than stated on the label (to WHO standards) |
|---|---|---|---|---|---|---|---|
| 1 | Off the shelf | Gel | Hand | Ethanol | 72 | 01±0.0 | Less (No) |
| 2 | Homemade | Gel | Hand | Ethanol | Not declared | 70±1.5 | (No) |
| 3 | Off the shelf | Gel | Hand | Ethanol | Not declared | 28±0.5 | (No) |
| 4 | Off the shelf | Gel | Hand | Ethanol | 70 | 09±0.0 | Less (No) |
| 5 | Off the shelf | Gel | Hand | Ethanol | Not declared | 29±0.5 | (No) |
| 6 | Off the shelf | Gel | Hand | Ethanol | 70 | 22±0.0 | Less (No) |
| 7 | Off the shelf | Gel | Hand | Ethanol | 68 | 40±0.5 | Less (No) |
| 8 | Off the shelf | Gel | Hand | Ethanol | 70 | 71±0.5 | (No) |
| 9 | Off the shelf | Gel | Hand | Ethanol | 70 | 21±0.5 | Less (No) |
| 10 | Off the shelf | Gel | Hand | Ethanol | Not declared | 30±0.0 | (No) |
| 11 | Off the shelf | Gel | Hand | Ethanol | 75% | 5.0±0.5 | Less (No) |
| 12 | Off the shelf | Spray | Hand/surface | Ethanol | 70+ | 90±0.5 | More |
| 13 | Off the shelf | Liquid | Hand/surface | Ethanol | 95 | 94±1.0 | **Yes** |
| 14 | Off the shelf | Liquid | Hand | Ethanol | 75 | 80±0.5 | More |
| 15 | Homemade | Liquid | Hand | Ethanol | 70 | 76±0.5 | More |
| 16 | Homemade | Liquid | Hand | Ethanol | 80 | 81±0.5 | NA |
| 17 | Homemade | Liquid | Hand/surface | Ethanol | 70 | 76±0.0 | More |
| 18 | Homemade | Liquid | Hand/surface | Ethanol | 70 | 70±1.0 | (No) |
| 19 | Homemade | Liquid | Hand | Ethanol | Not declared | 80±0.0 | NA |
| 20 | Homemade | Liquid | Hand | Ethanol | 70 | 65±1.0 | Less (No) |
| 21 | Homemade | Liquid | Hand | Ethanol | 80 | 81±1.0 | NA |
| 22 | Off the shelf | Liquid | Hand/surface | Ethanol | 70 | 75±0.5 | More |
| 23 | Off the shelf | Liquid | Hand | Ethanol | 70 | 76±0.5 | More |
| 24 | Off the shelf | Liquid | Hand/surface | Not declared | Not declared | 04±0.0 | (No) |
| 25 | Off the shelf | Liquid | Hand | Ethanol | 70 | 66±0.5 | Less (No) |
| 26 | Off the shelf | Liquid | Hand | Ethanol | 70 | 32±0.0 | Less (No) |
| 27 | Off the shelf | Liquid | Hand | Ethanol | 70 | 71±0.5 | (No) |
| 28 | Off the shelf | Liquid | Hand | Ethanol | Not declared | 09±0.0 | (No) |
| 29 | Off the shelf | Liquid | Hand | Not declared | Not declared | 29±0.0 | (No) |
| 30 | Off the shelf | Liquid | Hand/surface | Not declared | Not declared | 04±0.0 | (No) |
| 31 | Off the shelf | Liquid | Hand | Ethanol | Not declared | 48±0.5 | (No) |
| 32 | Off the shelf | Liquid | Hand | Ethanol | Not declared | 60±0.5 | (No) |
| 33 | Off the shelf | Liquid | Hand | Ethanol | 70 | 72±0.5 | (No) |
| 34 | Off the shelf | Liquid | Hand | Ethanol | 80 | 74±0.0 | Less |
| 35 | Off the shelf | Liquid | Hand | Ethanol | 84 | 76±0.0 | Less |
| 36 | Off the shelf | Liquid | Hand | Ethanol | Not declared | ND | (No) |
| 37 | Off the shelf | Liquid | Hand/surface | Isopropanol | 90+ | 99±0.5 | More |
| 38 | Off the shelf | Liquid | Hand | Ethanol | Not declared | 75±0.0 | NA |
| 39 | Off the shelf | Liquid | Hand | Ethanol | 80 | 22±0.0 | Less (No) |
| 40 | Off the shelf | Liquid | Hand | Ethanol | Not declared | 64±0.0 | (No) |
| 41 | Off the shelf | Liquid | Hand | Ethanol | Not declared | 78±0.0 | NA |
| 42 | Off the shelf | Liquid | Hand | Ethanol | Not declared | 56±0.0 | (No) |
| 43 | Off the shelf | Liquid | Hand/surface | Isopropanol | Not declared | 74±0.5 | NA |
| 44 | Off the shelf | Liquid | Hand | Ethanol | 70% | ND | (No) |
| 45 | Off the shelf | Liquid | Hand | Ethanol | 70% | 68±0.5 | (No) |
| 46 | Off the shelf | Liquid | Hand | Ethanol | Not declared | 66±0.5 | (No) |
| 47 | Off the shelf | Liquid | Hand | Ethanol | Not declared | 64±0.5 | (No) |
| 48 | Off the shelf | Liquid | Hand | Ethanol | Not declared | 60±0.0 | (No) |
| 49 | Off the shelf | Liquid | Hand/surface | Ethanol/Isobutanol | Not declared | 03/32* | (No) |
| 50 | Off the shelf | Liquid | Hand | Ethanol | 70 | 15±0.0 | Less (No) |

*NA, not applicable; ND, not detected*

**product with two different alcohol types*

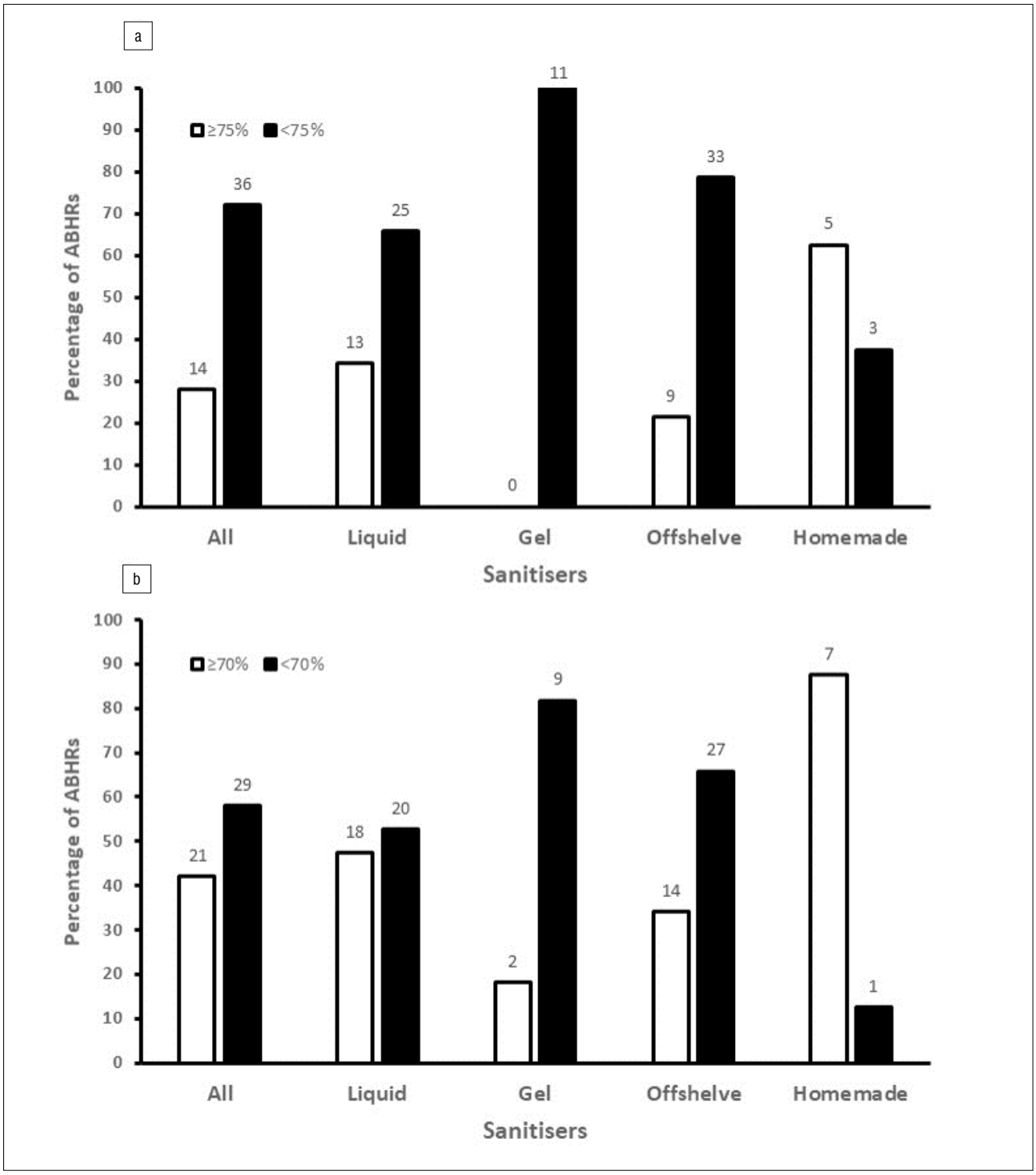



**Figure 3:** Proportion of products that contain at least (a) 75% and (b) 70% alcohol overall, based on formulation type and source.

It is also noteworthy that only WHO Formulations I, II, and variations thereof have been tested and found to be effective against SARS-CoV-2.[11] According to the WHO, an ABHR is 'an alcohol-containing preparation (liquid, gel or foam) designed for application to the hands to inactivate microorganisms and/or temporarily suppress their growth'[9]. Such preparations may contain one or more alcohols, and other active ingredients with excipients and humectants.[19] Considering this definition, and putting it in the context of the findings here, most of the products analysed do not qualify as ABHRs, but rather are cosmetics[15] because only products containing 70–95% alcohol have been shown to be able to denature the lipid and protein membranes of microorganisms[20]. It is only at 80% and 75%, respectively, for ethanol and isopropanol, that ABHRs are effective against emerging coronaviruses including SARS-CoV-2.[10,11] The findings here also corroborate earlier warnings by Korsten and de Bruinn[6] on the presence of fake hand sanitisers in the South African market and the need for South Africans to be protected against them.

In general, more liquid-based formulations were found to contain alcohols meeting the criteria for ABHRs in comparison to gel-based products. One reason for this is that liquid-based ABHR formulations are much easier to make in comparison to gel-based ones. For the former, ingredients are mixed based on volume per volume (v/v), while in the latter ingredients are measured based on weight per weight (w/w). Although

gel-based ABHRs are preferred due to their ease of use on hands and the presence of residual ingredients[21] such as perfumes. The efficacy of gel-based ABHRs is shrouded in controversy, with some studies showing no difference in their virucidal effect in comparison to those of liquid and foams[22] and some indicating that they are less efficient[23]. Aside from the alcohol contents and presence of residuals, the type and viscosity of the gel[3] used in the formulation also affects the delivery of the active ingredients in the required amounts, thus affecting virucidal activity. Even though making liquid-based formulations is easier, most of the off-the-shelf products analysed did not meet the criteria for ABHRs in comparison to homemade ABHRs that were formulated according to the WHO guidelines.[9] The production and sale of products that do not qualify as ABHRs in South Africa could be traced to several factors, including shortages of sanitising products experienced prior to the national lockdown resulting from the declaration of the state of disaster. These shortages and demands from consumers saw hoarding and an increase in prices. The increase in prices then led to the opening up of a market niche (valued in millions of rands) that made many companies, including non-chemical, pharmaceutical, and breweries, turn to producing ABHRs. In addition, ethanol – the most preferred alcohol in available products – is not cheap; hence cutting corners through reducing its composition by 10–20% translates into an increase in profit margins.

Correct labelling and declaration of contents for ABHRs are required in accordance with standards set by WHO and local regulatory agencies. Unfortunately, of the 50 products analysed, more than half did not declare their contents or made declarations that were inaccurate. In South Africa, the South African National Standards SANS 289 stipulates that labels on pre-packaged products should include its identity, name, place and business of the manufacturer, packer, distributor, importer/retailer and net quantity.[13] Likewise, SANS guide 490 clearly stipulated similar requirements on disinfectant alcohol-based hand rub products.[24] Practices involving non-declaration of contents and selling of products that are not of the required standard are infringing on consumer rights and are in contravention of the *South African Consumer Protection Act 68 of 2008*, which provides for fair, accessible products of high national and international standards. In addition, substandard ABHRs create a false sense of security about the efficaciousness of the products.

## Conclusion

A direct rapid and reproducible gas chromatography method for the determination of alcohols in hand-based sanitisers that can be used for the quality control of ABHRs was developed and optimised. Most of the commercial alcohol-based products sold and made available to consumers in public places are sub-standard and do not contain the required amount of alcohol to be classified as effective virucides, especially against SARS-CoV-2, the causative agent of COVID-19. The presence of products that do not qualify as ABHRs, and are not appropriately labelled, on the market as well as in public places, poses a great risk to consumers in the wake of preventative measures against COVID-19, more so because hand hygiene and disinfection remains one of the most (if not the only) effective measure for mitigating the spread of the disease available at this time. Thus, using sub-standard products exposes the population unknowingly to the virus by increasing the chances of transmission through contaminated surfaces. There is therefore a need to put in place quality control measures, especially at the manufacturing, wholesale and retail levels to ensure that the consumer gets good-quality ABHRs that qualify as virucides, and which are appropriately labelled. Added to this is the need to test ABHRs and any product sold as such for its virucidal effect to confirm its efficacy. For now, in the absence of appropriate quality control measures, preparing ABHRs using the WHO guide for local formulations remains a better alternative to purchasing off-the-shelf products that are mostly sub-standard.

## Acknowledgements

Thanks to volunteers who supplied hand sanitiser products and to colleagues for fruitful discussions on the idea to analyse hand sanitisers. Special thanks to Prof. Robin Crewe and Prof. Christian Pirk for reviewing an earlier version of the manuscript. Funding was provided for in part by the South African National Research Foundation grant for Y-rated researchers, SARChI Chair in Mathematical Models and Methods in Bio-engineering and Biosciences ($M^3B^2$) and the Alexander von Humboldt Foundation equipment grants. Views expressed in this article do not necessarily reflect the views of the funding bodies.

## Competing interests

I have no competing interests to declare.

## References

1. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med. 2020;383:727–733. https://doi.org/10.1056/nejmoa2001017
2. Center for Systems Science and Engineering (CSSE) Johns Hopkins University. Coronavirus COVID-19 global cases [webpage on the Internet]. c2020 [cited 2021 Oct 20]. Available from: https://coronavirus.jhu.edu/map.html
3. Berardi A, Perinelli DR, Merchant HA, Bisharat L, Bashetia IA, Bonacucina G, et al. Hand sanitisers amid CoViD-19: A critical review of alcohol-based products on the market and formulation approaches to respond to increasing demand. Int J Pherm. 2020;584, Art. #119431. https://doi.org/10.1016/j.ijpharm.2020.119431
4. Bomgardner MM, Mullin R, Scott A. Stepping up to the hand sanitizer shortage. C&EN Global Enterprise. 2020;98(11):12. https://doi.org/10.1021/cen-09811-buscon1
5. Grand View Research. Hand sanitiser market size, share and trend analysis report by product (gel, foam, liquid), by distribution channel (hypermarket & supermarket, drug store, speciality, store, online), by region, and segment forecasts, 2020–2027 [webpage on the Internet]. c2020 [cited 2020 Nov 23]. Available from: https://www.grandviewresearch.com/industry-analysis/hand-sanitizer-market
6. Korsten L, de Bruinn W. South Africans aren't being protected from fake sanitisers: what needs to be done. The Conversation. 2020 October 20. Available from: https://theconversation.com/south-africans-arent-being-protected-from-fake-sanitisers-what-needs-to-be-done-148128
7. Todd ECD, Michaels BS, Holah J, Smith D, Greig JD, Bartleson CA. Outbreaks where food workers have been implicated in the spread of foodborne disease. Part 10: Alcohol-based antiseptics for hand disinfection and a comparison of their effectiveness with soaps. J Food Prot. 2010;73:2128–2140. https://doi.org/104315/0362-028X-73.11.2128
8. Boyce JM, Pittet D. Guideline for hand hygiene in health-care settings: Recommendations of the Healthcare Infection Control Practices Advisory Committee and the HICPAC/SHEA/APIC/IDSA Hand Hygiene Task Force. Infect Control Hosp Epidemiol. 2002;23:S3–S40. https://doi.org/10.1086/503164
9. World Health Organization (WHO). Guide to local production: WHO-recommended handrub formulations. WHO/IER/PSP/2010.5 [webpage on the Internet]. c2010 [cited 2020 Nov 23]. Available from: https://www.who.int/publications/i/item/WHO-IER-PSP-2010.5
10. Siddharta A, Pfaender S, Vielle NJ, Dijkman R, Friesland M, Becker B, et al. Virucidal activity of World Health Organization-recommended formulations against enveloped viruses, including Zika, Ebola, and emerging coronaviruses. J Infect Dis. 2017;215(6):902–906. https://doi.org/10.1093/infdis/jix046
11. Kratzel A, Todt D, V'kovski P, Steiner S, Gultom M, Thao T, et al. Inactivation of Severe Acute Respiratory Syndrome Coronavirus 2 by WHO-recommended hand rub formulations and alcohols. Emerg Infect Dis. 2020;26(7):1592–1595. https://dx.doi.org/10.3201/eid2607.200915
12. Leslie RA, Zhou S, David RM. Inactivation of SARS-CoV-2 by commercially available alcohol-based hand sanitizers. Am J Infect Control. 2021;49(3):401–402. https://doi.org/10.1016/j.ajic.2020.08.020
13. South African Bureau of Standards (SABS). South African National Standard. Labelling requirements for prepackaged products (prepackages) and general requirements for the sale of goods subject to legal metrology control. SANS 298: 2016 Edition 1.06. Pretoria: South African Bureau of Standards (SABS) Standards Division; 2016.
14. Dhandapani R. Fast analysis of alcohol based hand sanitisers by gas chromatography. Phenomenex [webpage on the Internet]. c2020 [cited 2020 Dec 10]. Available from: https://phenomenex.blog/2020/03/31/alcohol-based-sanitizers/

15. Berardi A, Cenci-Goga B, Grispoldi L, Cossignani L, Perinelli DR. Analysis of commercial hand sanitisers amid COVID-19: Are we getting the products that we need? AAPS Pharm Sci Tech. 2020;21, Art. #286. https://doi.org/10.1208/s12249-020-01818-6

16. Fernando SF, Brito L, Pimetel MF, Leal LB. Determination of ethanol in gel hand sanitisers using mid and near infrared spectroscopy. J Braz Chem Soc. 2020;31(9):1759–1763. http://dx.doi.org/10.21577/0103-5053.20200115

17. Cartner T, Brand N, Tian K, Saud A, Carr T, Stapleton P, et al. Effect of different alcohols on stratum corneum kallikrein 5 and phospholipase A2 together with epidermal keratinocytes and skin irritation. Int J Cosmetic Sci. 2017;39:188–196. https://doi.org/10.1111/ics.12364

18. Tarka P, Gutkowska K, Nitsch-Osuch A. Assessment of tolerability and acceptability of an alcohol-based hand rub according to a WHO protocol and using apparatus tests. Antimicrob Resist Infect Control. 2019;8, Art. #191. https://doi.org/10.1186/s13756-019-0646-8

19. Gold NA, Mirza TM, Avva U. Alcohol sanitizer. StatPearls Publishing [webpage on the Internet]. c2020 [cited 2020 Nov 20]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK513254/

20. Jia Jing JL, Thong PY, Rajendran JC, McCarthy JR, Tharmalingam N, Madheswaran T. Hand sanitizers: A review on formulation aspects, adverse effects, and regulations. Int J Environ Res Public Health. 2020;17(9), Art. #3326. https://doi.org/10.3390/ijerph17093326

21. Kampf G, Kramer A, Suchomel M. Lack of sustained efficacy for alcohol-based surgical hand rubs containing 'residual active ingredients' according to EN 12791. J Hosp Infec. 2017;95:163–168. http://dx.doi.org/10.1016/j.jhin.2016.11.001

22. Larson EL, Cohen BV, Baxter KA. Analysis of alcohol-based hand sanitizer delivery systems: Efficacy of foam, gel, and wipes against influenza A (H1N1) virus on hands. Am J Infect Control. 2012;40:806–809. https://doi.org/10.1016/j.ajic.2011.10.016

23. Kramer A, Rudolph P, Kampf G, Pittet D. Limited efficacy of alcohol-based hand gels. Lancet. 2002;359:1489–1490. https://doi.org/10.1016/S0140-6736(02)08426-X

24. South African Bureau of Standards (SABS). South African National Standard disinfectant alcohol-based handrub. SANS 490: 2013 Edition 1.1 Amdt 1. Pretoria; South Africa Bureau of Standards (SABS) Standards Division; 2013. Available from: https://ctfa.co.za/wp-content/uploads/2020/04/SANS490_disinfectant-hand-rub-Standards_Development.pdf

**AUTHORS:**
Helga van der Merwe[1,2]
Suzanne J. Milton[1,3]
W. Richard J. Dean[1,3]
Tim G. O'Connor[1,4]
Joh R. Henschel[1,5]

**AFFILIATIONS:**
[1]South African Environmental Observation Network (SAEON), Kimberley, South Africa
[2]Plant Conservation Unit, Department of Biological Sciences, University of Cape Town, Cape Town, South Africa
[3]Percy FitzPatrick Institute of African Ornithology, Department of Biological Sciences, University of Cape Town, Cape Town, South Africa
[4]School of Animal, Plant and Environmental Sciences, University of the Witwatersrand, Johannesburg, South Africa
[5]Centre for Environmental Management, University of the Free State, Bloemfontein, South Africa

**CORRESPONDENCE TO:**
Helga van der Merwe

**EMAIL:**
h.vandermerwe@saeon.nrf.ac.za

**DATES:**
**Received:** 26 Mar. 2021
**Revised:** 21 June 2021
**Accepted:** 21 June 2021
**Published:** 29 Nov. 2021

**HOW TO CITE:**
Van der Merwe H, Milton SJ, Dean WRJ, O'Connor TG, Henschel JR. Developing an environmental research platform in the Karoo at the Square Kilometre Array. S Afr J Sci. 2021;117(11/12), Art. #10511. https://doi.org/10.17159/sajs.2021/10511

**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:**
drylands, ecosystem re-organisation, environmental change, long-term ecological research, Nama-Karoo

**FUNDING:**
National Research Foundation (South Africa)

# Developing an environmental research platform in the Karoo at the Square Kilometre Array

A part of the Square Kilometre Array (SKA) will be constructed in the northern Karoo of South Africa on approximately 135 000 ha of land. This land is formerly privately owned rangelands (farms) that were purchased by the South African National Research Foundation (NRF), on which the South African Radio Astronomy Observatory, as part of the global SKA project, will erect the SKA infrastructure. Additionally, a long-term environmental research programme will be established to investigate various dryland ecosystem components at a landscape scale. Livestock has been removed from the farms, and the area is now managed by the South African National Parks (SANParks) as the Meerkat National Park. The land-use and land cover changes present an unprecedented opportunity to study ecosystem dynamics. The property will be established as an NRF science park, incorporating an SKA research platform for radio astronomy and an environmental research platform of the South African Environmental Observation Network, with additional environmental research conducted by SANParks and their collaborators. We briefly describe current knowledge of the area's environment, and report on past and contemporary changes in this part of the Karoo. We present a conceptual model for the larger landscape which considers possible future land-use scenarios, the projected trajectories of change under these scenarios, and factors influencing these trajectories. These deliberations represent the foundation for future research in this landscape and the development of an environmental observation research platform in the Karoo at SKA.

**Significance:**

- We summarise an extensive environmental baseline report on the SKA property and surrounding areas.
- Withdrawal of livestock and other changes – such as clearing of alien invasive plants, reduced predator control and reduction in water-point maintenance – are expected to bring about changes in ecological processes and plant and animal communities.
- We present a conceptual model of scenarios to test possible future trajectories as a first step towards an earth system science research platform in the NRF science park.

## Introduction

Research sites that are owned by the state, and are secure and available for long-term environmental research, are few and far between. Such sites are disproportionately valuable[1-4] and are essential to provide data over time, 'linking biological patterns to environmental variability'[4]. In the drier parts of South Africa, such sites include Tierberg Karoo Research Centre (now Tierberg – Long-term Ecological Research) near Prince Albert[5], Grootfontein Agricultural Development Institute[6], Carnarvon Agricultural Research Station[7,8], and the Kalahari Meerkat study site near Van Zylsrus for investigation of evolutionary biology in animals[9]. Long-term sites elsewhere in the world have, for example, been very useful in identifying vegetation changes over several decades[10-12] and in providing data for overviews and modelling of desertification and changes in ecosystem function[13].

An opportunity to promote and expand long-term research, particularly in the drylands of the Karoo, has arisen through the establishment of the Square Kilometre Array (SKA), which is planned to be the largest radio telescope array in the world.[14,15] The core of the SKA, in the northern Nama-Karoo Biome (*sensu lato*)[16], offers an area in which the unusual combination of big-science astronomy concerning the universe[14,15] and earth system science concerning environmental changes on the ground[17], creates an opportunity for research on local climate, land-use and rangeland management, and social studies on rural and small-village societies and economies across an area of a million hectares[18-20]. The South African Environmental Observation Network (SAEON) (www.saeon.ac.za) has been in discussion with the South African Radio Astronomy Observatory (SARAO), the implementing agency of the SKA in South Africa, from an early stage and has produced a number of reports on vegetation and animals[21,22] in the area. Both SAEON and SARAO are National Research Infrastructure Platforms that reside within the South African National Research Foundation (NRF). The SAEON Arid Lands Node has a strong interest in long-term research sites, recognising that the value of long-term ecological research sites was constrained by the site-specificity of most studies, whereas drivers of change occurred beyond site boundaries.[23] The SKA, therefore, presents an opportunity to monitor and investigate key variables over many years using a sampling design that incorporates replication and land-type representativity.[23] This should improve the understanding of drivers of change across the arid lands in South Africa, what indicators describe the changes, what the baseline conditions of these indicators are, what changes are detected or predicted, where and why, and the implications of changes.[24]

Approximately 135 000 ha of land has been acquired by the NRF, where the highest concentration of the SKA radio astronomy infrastructure will be placed. This land is located within an area declared, in terms of the *Astronomy Geographic Advantage Act*[25], as the Karoo Central Astronomy Advantage Area. The NRF placed the SKA property, declared as the Meerkat National Park, under the protection of the South African National Parks (SANParks) in March 2020. Designating this property as an NRF science park, primarily for astronomy, enables

the SAEON Arid Lands Node to also establish a research platform for earth system sciences. Most of the Nama-Karoo is characterised by a repeated pattern of multi-year droughts[7], punctuated by boom times when the productivity of plants escalates across terrestrial and aquatic ecosystems, driving population irruptions of aquatic and terrestrial animals. These boom–bust dynamics may be affected by temperature increases, with new extremes of heat[26,27], and unusually long and intensive periods of drought[28]. The area is thus ideal for research into the dynamics of ecosystems in the Nama-Karoo.

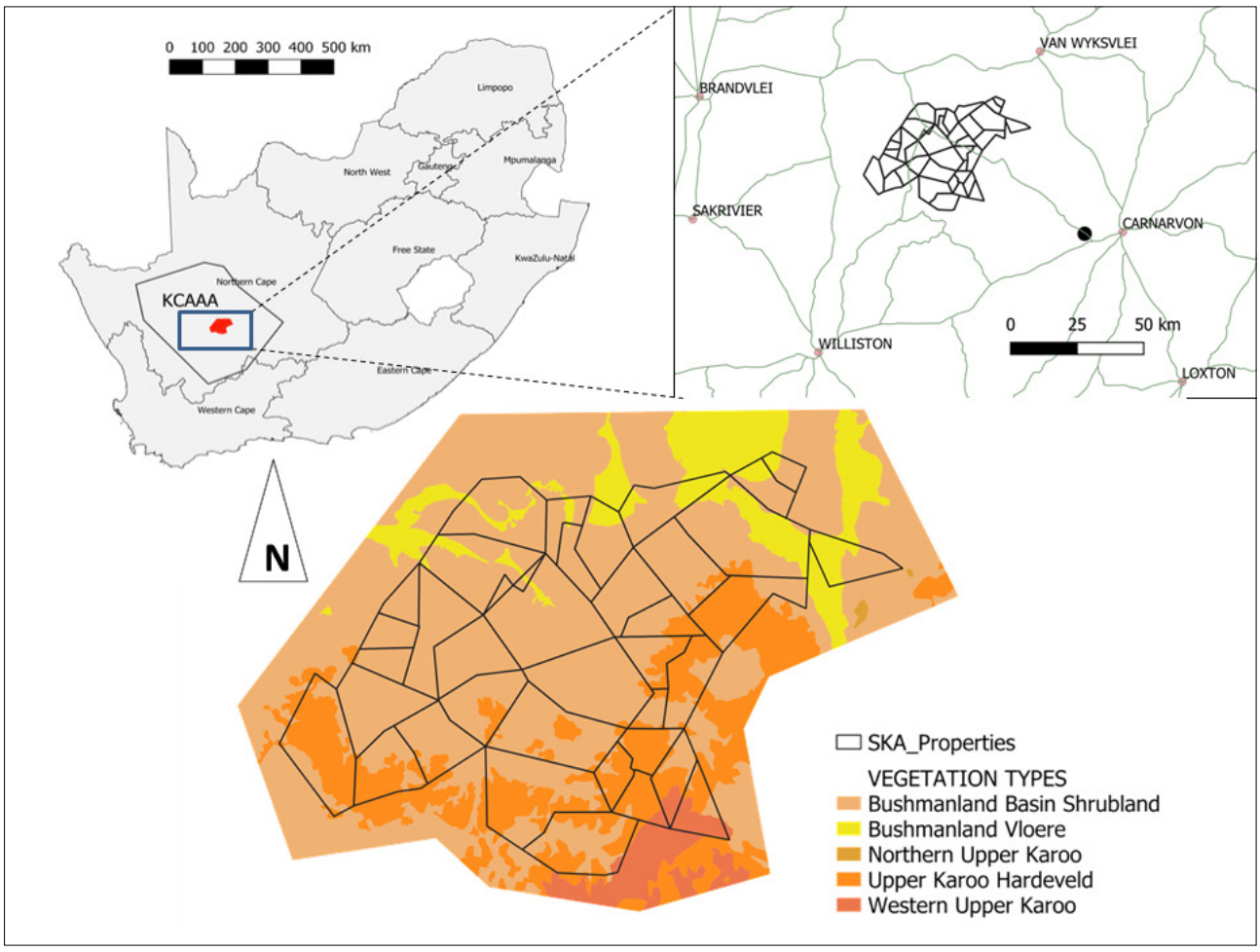
Under SARAO management, livestock (predominantly sheep at stocking rates of 30–39 ha/Large Stock Unit) was removed, resulting in a marked reduction in stocking pressure from June 2019. Accompanying the SKA infrastructure development[18,29] will be the removal of internal fencing, artificial water points and extensive stands of alien mesquite (*Prosopis* spp.) trees, and a gradual increase in wildlife numbers under the management of SANParks. Thus, the SKA and adjacent farmland are, in effect, a large-scale, long-term landscape-level experiment manipulating stocking rate, animal type, animal distribution and direct vegetation management. Resulting ecosystem re-organisation could include the gradual recovery of the system due to the release from livestock grazing and slow and returning wildlife or continuing ecosystem degradation due to increasing or sustained aridity.

Furthermore, past and current disturbances and the gradual natural or assisted rehabilitation of these areas could be investigated and inform restoration efforts that improve rangeland value in the region and on other drylands. Rehabilitation and/or restoration in dryland ecosystems is not well studied[30], and such research is mostly located in temperate areas[31]. The SKA, therefore, represents an unparalleled opportunity to discern the relative contribution of directional climate change and land use on ecosystem structure, functioning and diversity of a semi-arid area at the landscape scale. The reduction of the stocking rate to a fraction of what it has been for the past several decades and the transformation of the predominant land use – from rotational grazing systems under domestic stock by commercial farming to one of a continuous grazing system for large-scale wildlife management by a conservation authority – provides a unique opportunity to study a system returning, gradually, to conditions before livestock impacted the region.

Here, we present a position paper that summarises the current state of knowledge of the biota of the SKA (Meerkat National Park). It is essentially an advertisement for the opportunities for research that are offered by the study area. We encourage discussion about the most important research problems to be tackled within the scope of these opportunities, and invite collaborators with an interest in setting up environmental research projects at the SKA, subject to the requirements of SARAO and SANParks.

## Biogeophysical setting of the SKA environment

In South Africa, the core section of the SKA telescope is being developed in the Upper Karoo and Bushmanland region of the Northern Cape north of the Carnarvon Research Station between the villages of Brandvlei, Williston, Van Wyksvlei and Carnarvon and within the Nama-Karoo Biome (Figure 1). In planning to develop an environmental observation research platform, SAEON collated relevant information on the area in a comprehensive environmental baseline report[21], summarised in this section.




*KCAAA, Karoo Central Astronomy Advantage Area[25]*

**Figure 1:** Vegetation units[16] in the Square Kilometre Array (SKA) area. National Research Foundation property boundaries are indicated. The outer polygon is a 5–10 km buffer applied when mapping biophysical features of interest for comparing contrasting land uses.

### Climate

Most precipitation falls as rain, while hail, fog and snow are rarely recorded. Mean annual precipitation throughout the study area is less than 200 mm and increases with altitude from about 150 mm on the plains (960 m.a.s.l.) to 200 mm in the uplands (1100–1300 m.a.s.l.). Inter-annual variability in precipitation is high (CV = 44–60%). Above-average precipitation occurred during the mid-1970s, and short intermittent droughts occurred from 1934 to 1995. Currently, the area is experiencing a lengthy drought.[32] Rainfall cycles are apparent in the record at a period of 40–60 years[33] and for shorter periods[32].

Mean monthly temperature minima of 15 °C and maxima of 35 °C occur between December and February (summer) and minima of 5 °C and maxima of 20 °C between June and August (winter); however, extremes of 41 °C and -11 °C have been recorded in summer and winter, respectively. Mean humidity is below 50% for most of the year with a nadir in winter. Wind run peaks in the summer from November to February.[34] Potential monthly evaporation is, on average, 16-fold of monthly precipitation; the summer rate is 7-fold the winter rate. Evaporation from an A-Pan at Brandvlei has a mean of 5.5 mm/day (range 1–14 mm) and 165 mm/month with a mean annual evaporation (2002–2018) of 1992±82 mm.

Standardised Precipitation Evapotranspiration Index (SPEI) uses both precipitation and evapotranspiration to quantify drought.[28,35] The SPEI graph of the SKA area from January 1950 to August 2020 (Figure 2) shows the 12-month running mean compiled from SPEI-base data for 30.25°S, 21.75°E.[36] Drought indices below -2 signify severe drought. For the SKA area at the start of 2020, the index declined to -5, which is considered exceptionally extreme. Starting in 2014 and continuing until the end of the data set in August 2020, this drought is the longest and most intense recorded during the 70-year period of this record. Furthermore, a clear directional shift in SPEI is evident (Figure 2).

### Topography, geology and soils

Dolerite and Ecca shale flat-topped mesas are scattered across the plains. Ephemeral pans or 'vloere' occur in depressions in the lowest parts of the landscape. Soils of the elevated areas are skeletal, deeper alluvial soils occur on the plains and depressions, and colluvial deposits are present on the foot slopes. The area between the hills is covered with pebbly alluvium and sand and underlain by thinly layered shale of the Tierberg Formation shales.[37] The unconsolidated sand, silt and clay sediments are generally less than 2 m deep.[37] Calcrete is widespread in the landscape and occasionally quaternary, reddish aeolian sand can be found.[37]

### Hydrology

Hydrological features include an extensive network of ephemeral rivers, wetland depressions and overland drainage lines. Depressional wetlands occasionally hold shallow water over a large area after inundation. However, such events are rare, and most remain desiccated for many years or decades.[21,38] The surface flow of Karoo ephemeral rivers depends on groundwater discharge following precipitation at higher elevations, although there are few perennial springs and seeps.[39] The connection between the groundwater and surface water components remains poorly understood. The water table is generally shallow (12±5 m), and water quality is generally poor.[21]

### Evidence for climate change

Temperatures have risen throughout South Africa, with increases in warm extremes and decreases in cold spells being more substantial in the western regions.[26] In the SKA area, Van Wyksvlei (1939–2009) showed trends towards warmer extremes. The percentage of days per year when the maximum temperature exceeded the 90th percentile of the 1971–2000 base period increased at 2.26% per decade in the Northern Cape.[26] Van Wyksvlei and three other stations in the western central part of the country were the only stations for which data indicated a significant increase in the absolute annual maximum. Mean and maximum daily temperatures increased in the Northern Cape between 1960 and 2010, mostly after 2000.[27] There were no trends in annual or seasonal rainfall over this period, such that, even without a decline in rainfall, increased evapotranspiration resulting from higher temperatures was likely to have increased the frequency and severity of drought episodes, perhaps already evident in the region (Figure 2).[28]

### Biotic components

The SKA falls within the Nama-Karoo Biome.[16] Four Nama-Karoo vegetation units are represented, namely Western Upper Karoo, Upper Karoo Hardeveld and Northern Upper Karoo on the mountains and ridges, and Bushmanland Basin Shrubland on the lower altitude plains (Figure 1). Inland Azonal vegetation, comprising dry river beds and depressional wetlands collectively termed Bushmanland Vloere, intersperses and dissects these vegetation types.

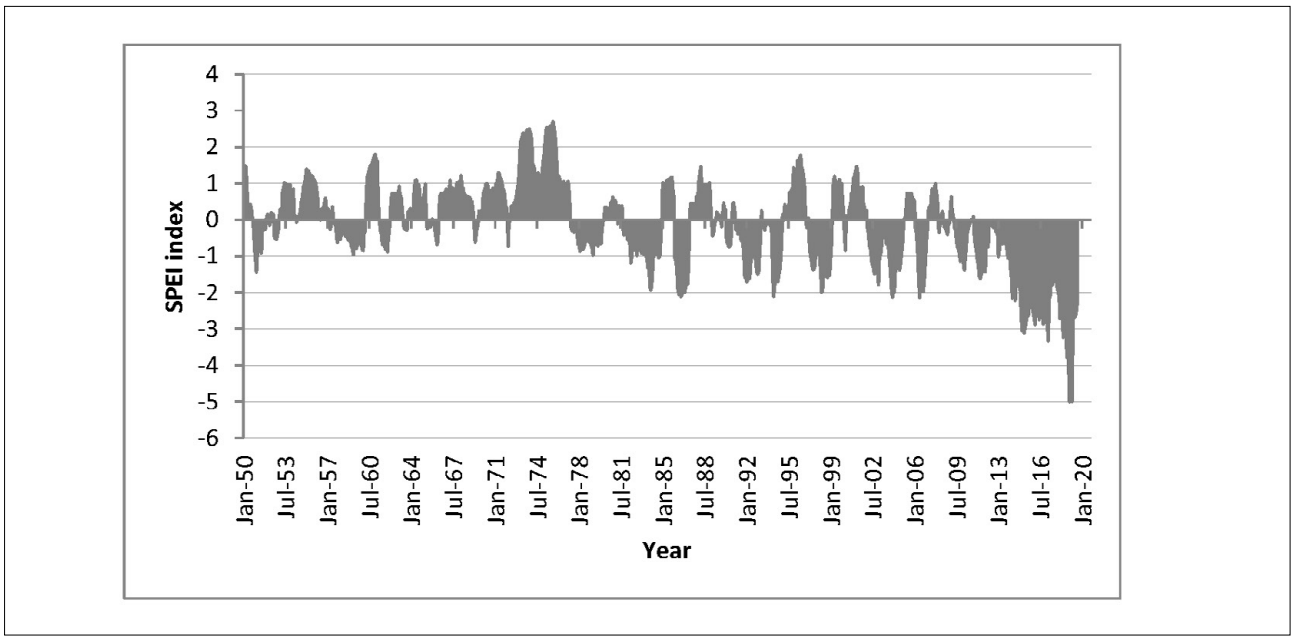



**Figure 2:** Standardised Precipitation and Evaporation Index (SPEI) for the Van Wyksvlei area from January 1950 to August 2020.[36]

The Milton et al.[21] baseline study reviewed the formal and grey literature, and database entries on the fauna and flora recorded from Quarter Degree Square 3021, and, to a lesser extent, its adjacent areas (Supplementary tables 1 and 2). The number of species recorded included 78 mammal, 142 bird, 56 reptile and 9 amphibian species. The avifauna, reptiles, amphibians and a few invertebrate groups such as darkling and dung beetles, and aquatic invertebrates are relatively well documented. The distributions of small mammals and bats require further investigation. Most invertebrate groups are poorly collected, and no records were found for some (e.g. molluscs, millipedes). The 10 most speciose floral families are Aizoaceae, Amaranthaceae, Apocynaceae, Asteraceae, Fabaceae, Hyacinthaceae, Iridaceae, Malvaceae, Poaceae, and Scrophulariaceae. However, ephemeral components such as fungi, geophytes and short-lived herbs are incompletely documented.

## Conditions for conducting environmental research at SKA

Apart from two recently published vegetation studies on the SKA property[22,40], the only other published ecological research from the vicinity are studies of vegetation from Carnarvon Research Station[7,8] and of invertebrates in nearby Bushmanland wetlands[38,41,42]. Most of the documentation of environmental conditions has been in government reports (e.g. Du Toit et al.[33]), or historical accounts by early travellers (e.g. Burchell[43]; Kirby[44]), or information extracted from early accounts (e.g. Skead et al.[45]; Boshoff et al.[46]).

The SKA offers an unprecedented opportunity to study ecosystem re-organisation across a landscape following cessation of sustained livestock grazing. Vegetation management treatments such as disturbance, restoration (natural and assisted), and alien vegetation removal are present in the same landscape. The potential value of this effort is enhanced because the SKA area is representative of a large part of the arid western interior of South Africa – an area experiencing increasing temperature and water stresses (Figure 2).[26] Findings can also help us understand natural drivers of change and how these impact social and economic elements in the region. State ownership of a large part of the study area should ensure uninterrupted long-term study and observation into the future that would contribute to the implementation of the Integrated Environmental Management Plan by the SARAO and the management of the Meerkat National Park by SANParks.[29]

### *Conceptual model of the possible impacts of land-use change in the SKA region*

#### Future land-use scenarios

The juxtaposition of the extensive livestock-free core area of the SKA with surrounding land still actively farmed with livestock allows testing of land-use scenarios at both small and large scales (Figure 3). Additionally, historical and recent disturbances have scarred the landscape, and the timespan or trajectory for recovery with or without intervention is unknown. Land degradation includes bush thickening by indigenous non-forage shrubs (e.g. *Rhigozum trichotomum*) and invasion by alien trees (*Prosopis* spp.)[21,22], which influence the landscape's biotic and abiotic components. Investigating recovery processes under optimal conditions in terms of stocking rate would facilitate ecological restoration at SKA that can inform Environmental Impact Assessments elsewhere and environmental management regulations in general.

Future land-use scenarios for the SKA and adjacent area may include: (Figure 3, scenario A) exclusion of livestock grazing with the low numbers of wildlife remaining on the SKA and slowly increasing through natural processes, (B) re-introduction of wildlife onto the property thereby increasing wildlife numbers and diversity more rapidly than would have occurred naturally, (C) adjacent rangeland with livestock, i.e. the status quo for most of the Nama-Karoo region, (D) some areas subjected to overgrazing by livestock and/or wildlife, (E) historically disturbed areas resulting from ploughing, cropping or other agricultural practices which fail to recover without active intervention, (F) the condition of past disturbed areas improving with active intervention, (G) recently disturbed areas remaining in a poor state or retrogressing without active intervention, or (H) recently disturbed areas stabilising or improving after active intervention.

#### Predicted trajectories of change

We predict that the scenarios mentioned above will affect private land and/or the SKA property as follows. Firstly, within constraints of low rainfall, the removal of livestock allows for a slow (>20 to 40 years) increase in vegetation cover and decrease in bare soils, accompanied by gradual changes in biotic community composition (e.g. Milton and Hoffman[47]; Van Rooyen et al.[48]). Secondly, re-introducing wildlife after removing livestock will follow the first trajectory unless wildlife numbers escalate (e.g. Van Rooyen et al.[48,49]). The third possible trajectory results from continued livestock grazing under sustainable stocking rates, driven by natural environmental fluctuations, as for the recent past. The fourth pathway is associated with overgrazing by livestock and/or wildlife causing decreasing vegetation cover, increasing soil exposure, and changing vegetation structure and species composition (e.g. Milton[50]; Milton and Hoffman[47]; Van Rooyen et al.[48,49]). Historically disturbed areas without intervention would follow a fifth pathway in which there is a slow increase in vegetation cover, decreasing exposed soils, and slow floral and faunal enrichment. A sixth trajectory accelerates and enhances the fifth with ecological restoration (e.g. Milton and Hoffman[47]). The seventh, with no active intervention at newly disturbed sites, leaves the area mostly bare, with low vegetation cover, depleted flora and fauna, and a state of degradation from which recovery is slow (e.g. Milton and Hoffman[47]). Recovery of vegetation cover and community composition at such sites would be quicker with active intervention, giving a final pathway.

#### Factors influencing trajectories of change

The land-use scenarios and the predicted trajectories of change are affected by numerous internal and external factors. These include the (1) historical, current and changing land use, (2) episodic events, (3) vegetation fluctuations, (4) faunal variations, (5) social influences, and (6) global change.

Historical disturbances (Figure 3, driver 1), especially overgrazing, ploughing, cropping and damming of rivers, remain visible in the Karoo vegetation for centuries (e.g. Milton[50]; Milton and Hoffman[47]; Milton et al.[21]). The combination of past and current land uses affects ecosystem changes (e.g. Milton[50]; Van der Merwe et al.[8]). Episodic events alter the physical landscape (Figure 3, driver 2). Exceptional wet periods drive vegetation flushes[16,51,52] and associated irruptions of animal populations and subsequent mass movements of birds[53], springbok[54], hyrax[55], locusts[56] and rapidly increase the abundance of branchiopods in pans[38]. Flash floods also cause degradation through erosion and affect geohydrology.[57] Droughts and erosive dust storms can result in massive dieback of vegetation and animal populations.[56,58]

Exceptionally wet periods promote large-scale recruitment of some plant species, with the seed bank and land management determining the resulting vegetation.[58,59] Similarly, reduced stocking rates allow plant species to grow, set seed and germinate, with recruits replacing palatable and unpalatable species at different frequencies[50,58,59] (Figure 3, driver 3).

In the Karoo, some faunal species are known for their population irruptions and crashes that follow bursts of primary productivity after rainfall, a pattern that is characteristic of usually dry areas.[16,52] Populations of hyrax, springbok, rodents and locusts, for example, irrupt during favourable conditions and then crash or move as food sources become scarce.[54-56] These population irruptions and crashes can have knock-on effects on predators, such as jackal (*Canis mesomelas*), caracal (*Caracal caracal*) and Verreaux's eagles (*Aquila verreauxii*) (Figure 3, driver 4). The amplitude of such population cycles could change following the absence of livestock and closure of artificial water points.

Social and economic pressures brought about by local, national and global factors affect employment and educational opportunities, state of health and well-being.[20] Additional to these influences, climatic events

such as long periods of drought may decrease rangeland condition, thereby compromising farming activities. This may lead to large-scale unemployment, with people moving to the surrounding villages, thereby placing additional pressure on natural resources such as fuelwood, wildlife and medicinal plants (Figure 3, driver 5). At a higher level, global change due to increasing $CO_2$ levels, temperatures, frequency and intensity of droughts or cold spells, and shifting seasons also has local effects (Figure 3, driver 6). These changes are not controlled at a local or national level, but knowledge of responses at these levels is necessary to inform adaptions and mitigation measures to continue working and living in the region.

## Development of an environmental research platform at the SKA

For long-term observation to contribute to improved understanding, surveillance of relevant drivers and appropriate response variables is essential. Future potential change is also more confidently addressed when there is a clearer understanding of historical change. To attain this aim, SAEON will monitor climate at a series of distributed sites at SKA using weather stations and other instrumentation, deployed according to regulations concerning radio frequency interference. A primary use of the weather data will be an effort to distinguish natural climatic variability from global increases. Climate monitoring is fundamental to interpret data of the abiotic and biotic environment at SKA and the adjacent farmland.

Vegetation can be monitored by establishing and surveying permanent plots at SKA and on adjacent farmland at different scales across vegetation units, land types and landforms[60], and ensuring data collection methods are constant through time[61]. This design will contribute to a landscape-scale (*sensu* Sparrow et al.[60]) network of study plots across different climatic zones in South Africa and enhance continental and global comparisons[62]. Similar methodology was developed for surveillance of Australian rangelands as part of the Australian Terrestrial Ecosystem Research Network.[62]

Responses to land use and the change in land use will be investigated by monitoring the impact of grazers (livestock and wildlife) on vegetation change, enhanced by studying the effect of variable stocking rates on animal types and distributions of individual animals. Biodiversity will be monitored in detail at a range of intensively studied sites and at less-intensely studied sites by monitoring plant and animal populations, community composition and population dynamics over time and across environmental gradients within the SKA and on adjacent farmland. In addition, ongoing surveillance during regular site visits to conduct monitoring is possibly the only means whereby infrequent events and their effects on the environment can become better understood.

Hydrological functioning will be investigated in order to better understand the hydrology of the Karoo and specifically its functioning in dryland pan systems. Prolonged and extreme drought and heat can elevate the significance of moisture availability in cooler months, driving fundamental changes in plant composition[63], for example, counterintuitively favouring C3 plants[64]. This emphasises the need to monitor seasonal shifts in rainfall and its effects. Historical shifts in plant community composition associated with past droughts could perhaps be detected by analysing plant remains in dung. The numerous hyrax middens at SKA represent a palaeoenvironmental archive[65], which can be actively calibrated with current measures of changes in plant community composition driven by climate change, drought, and land-use changes.

Global climate change and its influence on ecosystem structure, functioning and processes will impact not only the Meerkat National Park but also the surrounding sparsely populated agricultural land and the economy of the local villages that are vulnerable to environmental extremes.[66] The environmental dynamics of the area should be studied by integrating traditional disciplinary sciences in a transdisciplinary, holistic long-term socio-ecological research[67] approach.

## Desired outcomes of the environmental research platform at SKA

Land use and the resulting land-cover changes at SKA present unprecedented opportunities for arid lands research. The abrupt removal of livestock from a large area of rangeland and gradual system re-organisation under various land uses and stocking rates allow for the investigation of multiple treatments and trajectories of change in biotic and abiotic components. The changes can be compared to changes on adjacent rangelands still grazed by livestock. Long-term monitoring of changes in various components of the most arid ecosystem compared with less arid systems should provide insights and understanding of the contributions of global climate change superimposed on land-use change.

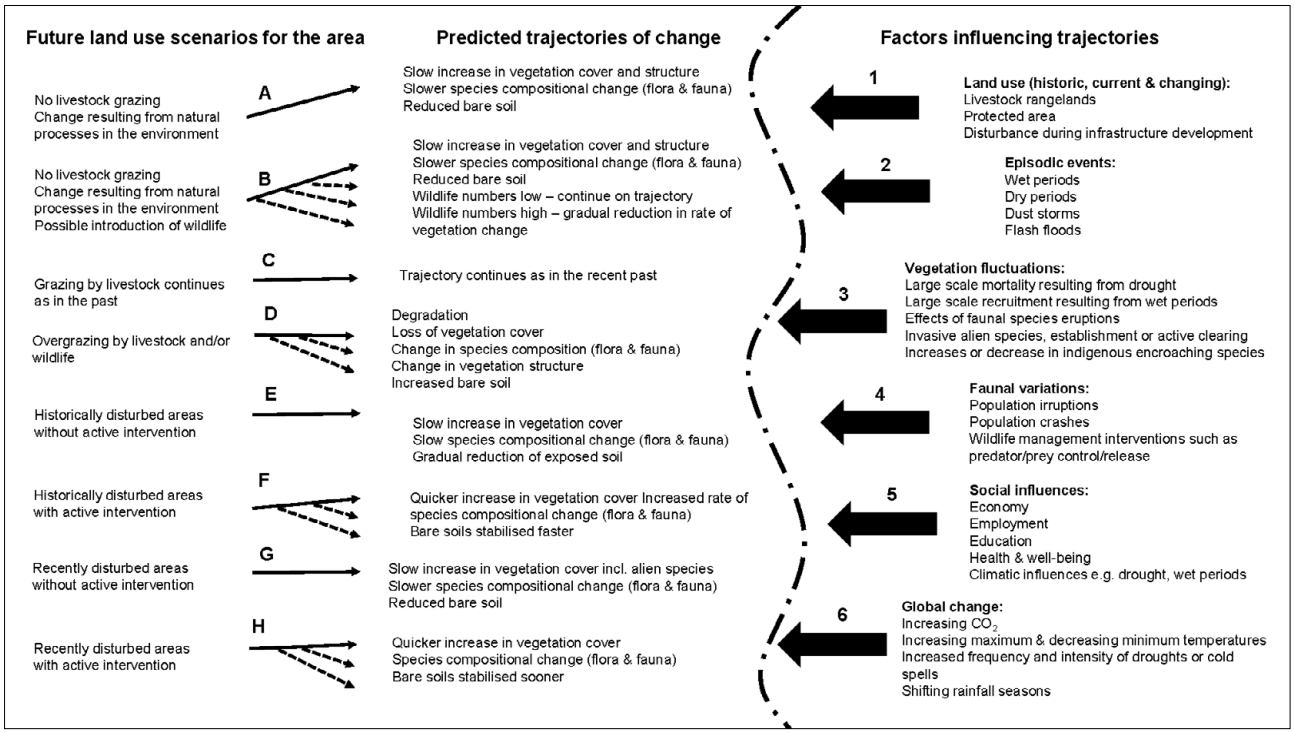



**Figure 3:** Conceptual model of future land-use scenarios, predicted trajectories of change and factors that could influence the predicted trajectories.

## Acknowledgements

The National Research Foundation (South Africa) provided some financial assistance, but the paper's content is solely the authors' responsibility. We thank Tracy Cheetham, Alice Pienaar-Marais, Dawie Fourie, Nomcebo Buthelezi and Gabriela Duncan of the South African Radio Astronomy Observatory (SARAO) for discussions and information concerning the SKA. South African National Parks (SANParks) is acknowledged for permission to work in Meerkat National Park and for discussions with Hugo Bezuidenhout, Hendrik Malgas, Nkabeng Maruping-Mzileni, Ernest Daemane and Abel Matsapula. Both SARAO and SANParks are thanked for their input into this manuscript.

## Competing interests

We have no competing interests to declare.

## Authors' contributions

All authors contributed to the conceptualisation of the article; S.J.M. and W.R.J.D. collated the baseline data; H.v.d.M. was responsible for writing the article with substantial input from all co-authors.

## References

1. Strayer D, Glitzenstein JS, Jones CG, Kolasa J, Likens GE, McDonnell MJ, et al. Long-term ecological studies: An illustrated account of their design, operation, and importance to ecology. Occasional Publication of the Institute of Ecosystem Studies. 1986;2.
2. Likens GE. Long-term studies in ecology. Approaches and alternatives. Berlin: Springer Verlag; 1989. https://doi.org/10.1007/978-1-4615-7358-6
3. Whitford WG. Jornada Validation Site Report. U.S. International Biological Program, Desert Biome. Final progress reports: Validation studies RM 77-4. Logan, UT: Utah State University;1979. p. 93–102. [cited 2021 Oct 14]. Available from: https://digitalcommons.usu.edu/cgi/viewcontent.cgi?article=1280&context=dbiome_memo
4. Hughes BB, Beas-Luna R, Barner AK, Brewitt K, Brumbaugh DR, Cerny-Chipman EB, et al. Long-term studies contribute disproportionately to ecology and policy. BioScience. 2017;67:271–281. https://doi.org/10.1093/biosci/biw185
5. Arena G, Van der Merwe H, Todd SW, Pauw MJ, Milton SJ, Dean WRJ, et al. Reflections, applications and future directions of long-term ecological research at Tierberg. Afr J Range Forage Sci. 2018;35(3–4):257–265. https://doi.org/10.2989/10220119.2018.1513072
6. Grootfontein Agricultural Development Institute [homepage on the Internet]. Available from: http://gadi.agric.za/
7. Harmse CJ, Gerber HJ. Effect of stocking density on vegetation and animal performance within the semi-arid shrublands of the Karoo, South Africa. Afr J Range Forage Sci. 2018;35:13–22. https://doi.org/10.2989/10220119.2018.1440629
8. Van der Merwe H, Du Toit JCO, Van den Berg L, O'Connor TG. Impact of sheep grazing intensity on vegetation at the Arid Karoo: Stocking Rate Trial after 27 years, Carnarvon, South Africa. J Arid Environ. 2018;155:36–45. https://doi.org/10.1016/j.jaridenv.2018.02.005
9. Clutton-Brock T, Sheldon BC. Individuals and populations: The role of long-term, individual-based studies of animals in ecology and evolutionary biology. Trends Ecol Evol. 2010;25:562–573. https://doi.org/10.1016/j.tree.2010.08.002
10. Buffington LC, Herbel CH. Vegetational changes on a semidesert grassland range from 1858 to 1963. Ecol Monogr. 1965;35:139–164. https://doi.org/10.2307/1948415
11. Hennessy JT, Gibbens RP, Tromble JM, Cardenas M. Vegetation changes from 1935 to 1980 in mesquite dunelands and former grasslands of southern New Mexico. J Range Manag. 1983;36:370–374. https://doi.org/10.2307/3898490
12. Schlesinger WH, Reynolds JF, Cunningham GL, Huenneke LF, Jarrell WM, Virginia RA, et al. Biological feedbacks in global desertification. Science. 1990;247:1043–1048. https://doi.org/10.1126/science.247.4946.1043
13. Reynolds JF, Grainger A, Stafford Smith DM, Bastina G, Garcia Barros L, Fernández RJ, et al. Scientific concepts for an integrated analysis of desertification. Land Degrad Dev. 2011;22:166–183. https://doi.org/10.1002/ldr.1104
14. Wild S. Searching African skies: The Square Kilometre Array and South Africa's quest to hear the songs of the stars. Johannesburg: Jacana Media; 2013.
15. Camilo F, Scholz P, Serylak M, Buchner S, Merryfield M, Kaspi VM, et al. Revival of the Magnetar PSR J1622–4950: Observations with MeerKAT, Parkes, XMM-Newton, Swift, Chandra, and NuSTAR. ApJ. 2018;856:180–191. https://doi.org/10.3847/1538-4357/aab35a
16. Mucina L, Rutherford MC, editors. The vegetation of South Africa, Lesotho and Swaziland. Pretoria: South African National Biodiversity Institute; 2006.
17. Zietsman L, editor. Observations on environmental change in South Africa. Stellenbosch: African SUN MeDIA; 2011.
18. Council for Scientific and Industrial Research (CSIR). Strategic environmental assessment for the South African mid-frequency array of SKA Phase 1. CSIR report CSIR/02100/EMS/ER/2016/15240/B. Pretoria: CSIR; 2016.
19. Gastrow M, Oppelt T. The Square Kilometre Array and local development mandates in the Karoo. J South Afr. 2019;45:711–728. https://doi.org/10.1080/03057070.2019.1642679
20. Walker C. Cosmopolitan Karoo: Land, space and place in the shadow of the Square Kilometre Array. J South Afr. 2019;45:641–662. https://doi.org/10.1080/03057070.2019.1645493
21. Milton SJ, Henschel JR, Van der Merwe H, Dean WRJ, Meyer-Milne E, Gerber H. Environmental baseline review of the core area and surrounds of the Square Kilometre Array (SKA). SAEON report and appendices. Pretoria: South African Environmental Observation Network (SAEON); 2021. https://doi.org/10.15493/saeon.arid.10000001
22. Van der Merwe H. Vegetation of the Square Kilometre Array (SKA), South Africa: A baseline to measure local and global change. S Afr J Bot. 2020;132:22–29. https://doi.org/10.1016/j.sajb.2020.03.025
23. Milton SJ, Dean WRJ, O'Connor TG, Mills AJ. Scaling up from site-based research to a national research and monitoring network: Lessons from Tierberg Karoo Research Centre and other design considerations. S Afr J Sci. 2007;103:311–317.
24. O'Connor TG. Understanding environmental change in complex systems: SAEON core science framework. Pretoria: South African Environmental Observation Network (SAEON); 2010.
25. Astronomy Geographic Advantage Act 21 of 2007, Republic of South Africa. South African Government Gazette. 2008;31157:1–56. https://www.gov.za/sites/default/files/gcis_document/201409/gg31157nn666apg1-30.pdf
26. Kruger AC, Sekele SS. Trends in extreme temperature indices in South Africa: 1962–2009. Int J Climatol. 2013;33:661–676. https://doi.org/10.1002/joc.3455
27. MacKellar N, New M, Jack C. Observed and modelled trends in rainfall and temperature for South Africa: 1960–2010. S Afr J Sci. 2014;110(7/8), Art. #2013-0353, 13 pages. https://doi.org/10.1590/sajs.2014/20130353
28. Vicente-Serrano SM, Beguerias S, Lopez-Moreno JI. A multiscalar drought index sensitive to global warming: The Standardised Precipitation Evapotranspiration Index. J Clim. 2010;23:1696–1718. https://doi.org/10.1175/2009JCLI2909.1
29. South African Radio Astronomy Observatory, National Research Foundation. Integrated Environmental Management Plan (IEMP) for SKA Phase 1 mid-frequency array (SKA1_MID) in South Africa [document on the Internet]. No date [cited 2021 Jun 21]. Available from: https://www.sarao.ac.za/about/strategic-environmental-assessment/
30. Aronson J, Milton SJ, Blignaut JN. Restoring natural capital: Science, business, and practice. Washington / London: Island Press / Covelo; 2007.
31. Shriver RK, Andrews CM, Pilliod DS, Arkle RS, Welty JL, Germino MJ, et al. Adapting management to a changing world: Warm temperatures, dry soil, and interannual variability limit restoration success of a dominant woody shrub in temperate drylands. Glob Chang Biol. 2018:1–11. https://doi.org/10.1111/gcb.14374

32. Harmse CJ, Du Toit JCO, Swanepoel A, Gerber HJ. Trend analysis of long-term rainfall data in the Upper Karoo of South Africa. Trans R Soc S Afr. 2021;76(1):1–12. https://doi.org/10.1080/0035919X.2020.1834467

33. Du Toit JCO, Van den Berg L, O'Connor TG. A summary of rainfall at the Carnarvon Experiment Station, 1931-2013. Grootfontein Agric. 2015;15:27–35.

34. Meyer TC. Weidingskapsiteitstudies op veld in die dorre Karoo [Grazing capacity studies on veld in the arid Karoo] [thesis]. Bloemfontein: University of the Free State; 1992. Afrikaans.

35. Beguería S, Vicente-Serrano SM, Angulo MA. Multiscalar global drought dataset: The SPEIbase: A new gridded product for the analysis of drought variability and impacts. Bull Am Meteorol Soc. 2010;91:1351–1356. https://doi.org/10.1175/2010bams2988.1

36. Global SPEI database [webpage on the Internet]. No date [cited 2020 Oct 12]. Available from: http://spei.csic.es/database.html

37. Cronwright MS, Bosch PJA, Forbes C, Cole J. Assessment of the geological and geotechnical constraints and mitigation measures which may be required for the location of the Square Kilometre Array core site and the Karoo array telescope on the Karoo and Namaqualand sites. Internal report No.: 2005–0121. Pretoria: Council for Geosciences; 2005.

38. Meyer-Milne E, Mlambo MC, Rogers DC. Distribution of clam shrimps (Crustacea: Laevicaudata and Spinicaudata) in South Africa, with new records from the Northern Cape Province. Zool Stud. 2020;59:39. https://doi.org/10.6620/ZS.2020.59-39

39. Snaddon K, Ollis D, Ngobela T, Kirkwood D. Aquatic ecosystems assessment of the SKA Phase 1 in South Africa. Pretoria: CSIR; 2017. Available from: http://www.skaphase1.csir.co.za/wp-content/uploads/2017/01/Aquatic-Assessment.pdf

40. Van der Merwe H, Geldenhuys C. Proposed long-term monitoring protocol and applications for *Aloidendron dichotomum* populations. S Afr J Bot. 2017;109:253–262. https://doi.org/10.1016/j.sajb.2017.01.008

41. Hamer M, Rayner NA. A note on the unusual Crustacean community of a temporary pool in the Northern Cape. S Afr J Aquatic Sci. 1996;22(1/2):100–104. https://doi.org/10.1080/10183469.1996.9631376

42. Anderson MD. The status of flamingos in the Northern Cape Province, South Africa. Ostrich. 2000;71:425–437. https://doi.org/10.1080/00306525.2000.9639859

43. Burchell WJ. Travels in the interior of southern Africa, Vol. 1. London: The Batchworth Press; 1822. Reprinted 1953. https://doi.org/10.5962/bhl.title.100911

44. Kirby PR, editor. The diary of Dr Andrew Smith, Director of the 'Expedition for Exploring Central Africa', 1834-1836. Publications No. 20 & 21, 1939-1940. Cape Town: Van Riebeeck Society; 1940.

45. Skead CJ, Boshoff A, Kerley G, Lloyd P. Historical incidence of the larger land mammals in the broader western and northern Cape. 2nd ed. Port Elizabeth: Centre for African Conservation Ecology, Nelson Mandela University; 2011.

46. Boshoff A, Landman M, Kerley G. Filling the gaps on the maps: Historical distribution patterns of some larger mammals in part of southern Africa. Trans Roy Soc S Afr. 2016;71:23–87. https://doi.org/10.1080/0035919x.2015.1084066

47. Milton SJ, Hoffman M.T. The application of state-and-transition models to rangeland research and management in arid succulent and semi-arid grassy Karoo, South Africa. Afr J Range Forage Sci. 1994;11:18–26. https://doi.org/10.1080/10220119.1994.9638349

48. Van Rooyen MW, Le Roux A, Geldenhuys C, Van Rooyen N, Broodryk NL, Van der Merwe, H. Long-term vegetation dynamics (40 yr) in the Succulent Karoo, South Africa: Effects of rainfall and grazing. Appl Veg Sci. 2015;18:311–322. https://doi.org/10.1111/avsc.12150

49. Van Rooyen MW, Le Roux A, Van der Merwe H, Van Rooyen N, Geldenhuys C. Long-term vegetation change (>20 years) in the plains habitat on the Goegap Nature Reserve, Succulent Karoo, South Africa. Afr J Range Forage Sci. 2018;35(3–4):289–302. https://doi.org/10.2989/10220119.2018.1498802

50. Milton SJ. Growth, flowering and recruitment of shrubs in grazed and in protected rangeland in the arid Karoo, South Africa. Vegetatio. 1994;111:17–27.

51. Lloyd JW. Phytosociology of the Vaalputs radioactive waste disposal site, Bushmanland, South Africa. S Afr J Sci. 1989;55:372–382. https://doi.org/10.1016/s0254-6299(16)31191-7

52. Milton SJ, Dean WRJ. Disturbance, drought and dynamics of desert dune grassland, South Africa. Plant Ecol. 2000;150:37–51. https://doi.org/10.1023/A:1026585211708

53. Dean WRJ. Nomadic desert birds. Berlin: Springer; 2004. https://doi.org/10.1007/978-3-662-08984-2

54. Roche CJ. Ornaments of the desert – Springbok treks in the Cape Colony, 1774-1908 [thesis]. Cape Town: University of Cape Town; 2004.

55. Davies RAG. Black eagle *Aquila verreauxii* predation on rock hyrax in the Karoo [PhD thesis]. Pretoria: University of Pretoria; 2000.

56. Henschel JR. Locust times – monitoring populations and outbreak controls in relation to Karoo natural capital. Trans R Soc S Afr. 2015;70:135–143. https://doi.org/10.1080/0035919X.2015.1046974

57. Vickery KJ, Eckardt FD, Bryant RG. A sub-basin scale dust plume source frequency inventory for southern Africa, 2005-2008. Geophys Res Lett. 2013;40:5274–5279. https://doi.org/10.1002/grl.50968

58. Van der Merwe H, Milton SJ. Testing the Wiegand–Milton model: A long-term experiment to understand mechanisms driving vegetation dynamics in arid shrublands. Austral Ecol. 2019;44:49–59. https://doi.org/10.1111/aec.12651

59. Milton SJ, Wiegand T. How grazing turns rare seedling recruitment events to non-events in arid environments. In: Breckle S-W., Veste M, Wucherer W, editors. Sustainable land-use in deserts. Heidelberg: Springer; 2001. https://doi.org/10.1007/978-3-642-59560-8_20

60. Sparrow BD, Edwards W, Munroe SEM, Wardle GM, Guerin GR, Bastin JF, et al. Effective ecosystem monitoring requires a multi-scaled approach. Biol Rev. 2020;95(6):1706–1719. https://doi.org/10.1111/brv.12636

61. Lindenmayer DB, Burns EL, Tennant P, Dickman CR, Green PT, Keith DA, et al. Contemplating the future: acting now on long-term monitoring to answer 2050's questions. Austral Ecol. 2015;40(3):213–224. https://doi.org/10.1111/aec.12207

62. Sparrow BD, Foulkes JN, Wardle GM, Leitch EJ, Caddy-Retalic S, Van Leeuwen SJ, et al. A vegetation and soil survey method for surveillance monitoring of rangeland environments. Front Ecol Evol. 2020;8:15–17. https://doi.org/10.20944/preprints201911.0178.v1

63. Duniway MC, Petrie MD, Peters DPC, Anderson JP, Crossland K, Herrick JE. Soil water dynamics at 15 locations distributed across a desert landscape: insights from a 27-yr dataset. Ecosphere. 2020;9(7):1–36. https://doi.org/10.1002/ecs2.2335

64. Knapp AK, Chen A, Griffin-Nolan RJ, Baur LE, Carroll CJW, Gray JE, et al. Resolving the Dust Bowl paradox of grassland responses to extreme drought. Proc Natl Acad Sci USA. 2020;117:22249–22255. https://doi.org/10.1073/pnas.1922030117

65. Chase BM, Scott L, Meadows ME, Gil-Romera G, Boom A, Carr AS, et al. Rock hyrax middens: A palaeoenvironmental archive for southern African drylands. Quat Sci Rev. 2012;56:107–125. https://doi.org/10.1016/j.quascirev.2012.08.018

66. Toerien D. Productive knowledge, poverty and the entrepreneurial challenges of South African towns. S Afr J Sci. 2018;114(11/12), Art. #4765, 8 pages. https://doi.org/10.17159/sajs.2018/4765

67. Dirnböck T, Haase P, Mirtl M, Pauw J, Templer PH. Contemporary International Long-Term Ecological Research (ILTER) – from biogeosciences to socio-ecology and biodiversity research. Reg Environ Change. 2019;19:309–311. https://doi.org/10.1007/s10113-018-1445-0

**AUTHORS:**
Hlekani V. Mbhatsani[1]
Ngoako S. Mabapa[1]
Tambe B. Ayuk[2]
Tshifhiwa C. Mandiwana[1]
Lindelani F. Mushaphi[1]
Merriam Mohlala[3]
Xikombiso G. Mbhenyane[2]

**AFFILIATIONS:**
[1]Department of Nutrition, University of Venda, Thohoyandou, South Africa
[2]Division of Human Nutrition, Stellenbosch University, Stellenbosch, South Africa
[3]Centre for Biokinetics, Recreation and Sport Science, University of Venda, Thohoyandou, South Africa

**CORRESPONDENCE TO:**
Xikombiso Mbhenyane

**EMAIL:**
xgm@sun.ac.za

**DATES:**
**Received:** 01 Sep. 2020
**Revised:** 19 Feb. 2021
**Accepted:** 23 July 2021
**Published:** 29 Nov. 2021

**HOW TO CITE:**
Mbhatsani HV, Mabapa NS, Ayuk TB, Mandiwana TC, Mushaphi LF, Mohlala M, et al. Food security and related health risk among adults in the Limpopo Province of South Africa. S Afr J Sci. 2021;117(11/12), Art. #8848. https://doi.org/10.17159/sajs.2021/8848

**ARTICLE INCLUDES:**
☒ Peer review
☐ Supplementary material

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

**EDITORS:**
Pascal Bessong
Sandiswa Mbewana

**KEYWORDS:**
household food insecurity, obesity, hypertension, health risk determinants, adults, South Africa

**FUNDING:**
University of Venda

# Food security and related health risk among adults in the Limpopo Province of South Africa

Food insecurity, obesity and hypertension remain major public health issues related to nutrition in South Africa. The purpose of this study was to determine household food security and the health risk of the adult population in the Limpopo Province using cross-sectional designs. A stratified random sampling method was used to recruit adults aged 18 to 65 years in the Limpopo Province of South Africa. Data were collected using a validated, structured questionnaire. All data were analysed using SPSS version 25.0. The study included 640 participants with an average age of 36.2±17.6 years and a household size of five persons; 74.5% of participants fell in the low monthly income bracket (≤ZAR3000). The mean dietary diversity score was 3.99 (CI: 2.79–5.19). The prevalence of food insecurity was 31.3%, obesity 35.2% and hypertension 32.3%. Being a woman, older and married significantly positively influenced obesity and hypertension. Also, a healthy eating lifestyle such as high dietary diversity was found to positively influence obesity status, while daily eating of fruit and vegetables positively significantly influenced the hypertension status of participants ($p<0.05$). Food insecurity, obesity and hypertension rates remain high among adults in the Limpopo Province of South Africa with consumption of a diet low in dietary variety. Aged and married women were more likely to be obese and hypertensive, while daily fruit and vegetable intake were found to be a protective factor. Educational and nutritional intervention should be designed and geared towards promoting fruit and vegetable intake in the community.

**Significance:**

- Households had a diet low in dietary variety, with dietary diversity revealed as a determinant of health risk.
- Dietary diversity is inversely correlated with household food security.
- The findings also identified household determinants of obesity and hypertension, which are major public health issues in South Africa.

## Introduction

Culture is learned behaviour common to a given society and shapes behaviour and consciousness of that particular society from one generation to another.[1] Food intake is heavily influenced by social, economic, political and cultural processes. Food habits are said to be a sub-system of a material culture, social structure and ideology that includes values, attitudes and norms.[2] Current eating habits of Africans reflect multicultural influences and inadequate knowledge of the indigenous eating culture.[3] Different cultures have different customs, food beliefs and taboos that influence their eating patterns.[4]

South Africa is a developing country with a heterogeneous socio-economic and multicultural society. The current eating patterns of black rural Africans indicate three porridge meals daily which are high in carbohydrates, moderate in protein and low in fat. Meat is consumed more regularly than in the past. The traditional diet has been abandoned due to the evolving food system. A nutrition transition has been noticed and it is the key contributor to the obesity epidemic, characterised by a change from a traditional diet to a Western-style diet.[5] Large shifts have occurred in both diet and physical activity patterns, particularly in the last one to two decades of the 20th century.[6]

South Africa has a triple burden of disease, exacerbated by the HIV/Aids epidemic. A consequence of these competing priorities is that there is little recognition of the magnitude of the burden of non-communicable diseases in South Africa. A demographic surveillance area study conducted by the Africa Centre for Health and Population Studies in rural KwaZulu-Natal reported that the prevalence of obesity and hypertension among adults 15–50 years old was 32% and 24% in 2004, respectively.[7] Also, from 1997 to 2004 in South Africa, 195 people died every day due to some form of heart and blood vessel disease (cardiovascular diseases). It is estimated that one in four South Africans between the ages of 15 and 64 years suffer from high blood pressure – one of the leading causes of heart attacks, strokes, kidney failure and premature death.[8] Furthermore, 39% of the reported mortality in 2010 was due to non-communicable diseases, increasing considerably to 51% of deaths in 2013.[8]

Many studies have reported that food insecurity is negatively associated with nutritional status and health outcomes such as acute and chronic infections, iron-deficiency anaemia, developmental and mental health problems.[9-13] In addition, a number of studies have reported an association between food insecurity and overweight/obesity among adults.[14,15] However, a limited number of studies have been conducted to evaluate the association between food insecurity and chronic diseases among the adult population. Most of the studies made use of self-reported chronic diseases measures to evaluate the association with food insecurity, without confirmation through objective assessment of the reported diseases.[14,16]

The purpose of this study was, therefore, to determine household food security and health risk of the adult population in the Limpopo Province. The study explored different factors combined that may be contributing to the health risk and nutritional status of the study population.

## Methods

The design was a cross-sectional, correlational study that surveyed households in the Limpopo Province of South Africa. The sample size was calculated from a population of 1 537 483 using Slovin's formula with 95% confidence level and 5% margin of error, which yielded a sample size of 385 from three of the six districts. A higher number was targeted to control for attrition and withdrawals. Stratified random sampling was used to recruit 640 adults (men and women) aged between 18 and 65 years. Data were collected using a validated questionnaire. The questionnaire consisted of five main sections namely: demographic data (gender, age, marital status, level of education, family size and household income), dietary patterns, household food security (determined using the 24-hour qualitative recall, hunger scale and food inventory), physical activity patterns and anthropometric measurements, specifically body mass index (BMI)/obesity and blood pressure.

The 24-hour recall, hunger scale and food inventory were used to assess consumption and availability of foods in a household. The prevalence of hunger was measured using a hunger scale questionnaire adapted from the World Health Organization (WHO) standardised tool used in the South African National Food Consumption Survey of 1999.[17] The physical activity patterns were assessed by asking questions that addressed sedentary lifestyle and physical exercise practice.

BMI and blood pressure were evaluated using both self-reported and clinical measurements. Weight and height were measured thrice using a calibrated electronic scale and stadiometer, respectively, and an average was computed. BMI was calculated from the weight and height measurements (weight in kg/height in metres squared). Systolic and diastolic blood pressures were measured in triplicate in a seated position with 5-min intervals between repeat measurements. Clinical evidence of hypertension was established to be either a measured systolic blood pressure (SBP) >140 mmHg and/or diastolic blood pressure (DBP) >90 mmHg. The self-reported use of antihypertensive medication was also considered.

### *Ethical consideration*

This study was approved by the higher degree and ethics committee of the University of Venda (SHS/08/NUT/003). Permission was also obtained from the chiefs or local councillors. The participants signed a consent form after the study purpose was explained to them. The research adhered to the principles of the Declaration of Helsinki.

### *Data analysis*

Data were analysed using SPSS version 25.0. Descriptive statistics were used to depict overall food security, obesity and hypertension status. Chi-square tests were used to examine associations between these outcome variables (obesity and hypertension status) and independent variables (socio-demographic characteristics, food insecurity and physical activity). The relationship between the participants' BMI and systolic pressure, dietary diversity score (DDS) and obesity and other variables were investigated using Pearson's correlation coefficient. The association between participants' socio-demographic parameters and DDS of ≤4 food groups was done using logistic regression analysis. Probability ($p$) values less than 0.05 were considered statistically significant.

## Results

### *Socio-demographic profile*

The study included 640 individuals with an average age of 36.2±17.6 years. The average age of men was 32.5 years while the average age of the women was 37.2 years. A predominant proportion (74.5%) of the sample had low monthly income (≤ZAR3000), and only 25.2% had a monthly income of above ZAR3000. An average of five persons lived in the surveyed households. The majority (81.5%) of the participants had secondary level of education while just 18.5% did not. Less than two-thirds (61.6%) of the participants were single and 38.4% were married (Table 1).

### *Health risk*

Regarding the physical activity level of the participants, 85.1% were involved in physical activity (self-reported), with household chores (19.5%) and walking (18.2%) being the most frequently reported. Three-quarters (75.5%) of the participants liked watching TV, with 58.5% indicating that they watch for more than 2.5 h at a time.

A higher proportion of women (43.6%) than men (25.6%) were obese or overweight (Figure 1). The mean systolic blood pressure was 121.1±20.3 mmHg, while diastolic blood pressure was 78.0±15.8 mmHg. The study revealed that 32.3% of participants showed clinical evidence of hypertension; of these clinically hypertensive participants, 80.7% were unaware of their condition and therefore were not receiving any treatment. Further analysis also showed that 61.3% of these hypertensive participants were over 60 years old.

### *Dietary patterns*

Dietary pattern analysis showed that the majority (92.2%) of participants mostly consumed food prepared at home and did so alone (79.2%). Approximately two-thirds (67.2%) of the participants consumed fruit and vegetables two to three times a week, while 16% reported that they consumed fruit and vegetables daily. In addition, 39.2% reported that they consumed fruit as snacks, while 35.8% consumed NikNaks™ (a local brand of puffed maize) as a snack. Regarding the number of meals eaten per day, 46.8% had two meals per day as compared to 44.3% who had three meals per day (Figure 2). The 24-hour recall revealed that mealie meal, salt, sugar, tea, bread and meat were the most consumed foods in the surveyed households. In addition, the household food inventory revealed that most of the households had mealie meal (32.3%), white bread (38.3%), brown bread (24.2%), corn flakes (17.5%), Rice Krispies® (14.3%) and Weet-Bix™ (8.0%) on the day of the survey.

### *Dietary diversity of the households*

The household dietary diversity of participants was assessed and categorised into three groups: low, medium and high dietary diversity. The findings show that 49.9% of the participants' households had medium dietary diversity, 25.6% low dietary diversity and 24.5% high dietary diversity (more than six food groups). An average household consumed four food groups a day. The food groups consumed by at least 50% of households were also assessed and findings revealed that cereal was predominately consumed at all the different levels of dietary diversity. Cereal was the main food group consumed by those with the lowest dietary diversity. The medium dietary diversity group mostly consumed cereal and dark green leafy vegetables. The households classified as high dietary diversity households mostly consumed cereals, vitamin A rich vegetables and tubers, other vegetables, and oils and fats (Table 2). The mean DDS was 3.99 (confidence interval (CI): 2.79–5.19).

The household dietary diversity was significantly associated with age group and income of the household after adjusting for other factors. Participants ≤35 years and 36–60 years were 13% (OR: 0.13; 95% CI: 0.03–0.63) and 11% (OR: 0.11; 95% CI: 0.02–0.58) less likely to have a DDS<4, respectively, than the age group of 60 years and above. Participants in households with a monthly income of less than ZAR3000 were twice (OR: 2.44; 95% CI: 1.12–5.34) as likely to have a DDS<4 food groups (Table 3).

### *Food security*

Food insecurity was higher in the present study, while obesity and hypertension were lower than the national rates. The findings reveal that 32.5% of the households surveyed were food secure, 36.3% of participants were at risk of hunger and 31.3% experienced hunger (Figure 3). The present findings align with the previous rates of food security found by the Limpopo provincial and national South African National Health and Nutrition Examination Survey published in 2013.[18] The Survey found a significant inverse correlation between household food insecurity and dietary diversity ($r$=-0.14, $p$=0.043).

**Table 1:** Participants' characteristics, obesity and hypertension status

| Demographic and dietary characteristics of participants | *n*=640 | Obese | Not obese | *p*-value | *n*=640 | SBP>140 or DBP>90 (mmHg) | Normal blood pressure | *p*-value |
|---|---|---|---|---|---|---|---|---|
| **Gender** | | | | **0.001** | | | | **0.035** |
| Male | 137 | 11 (9.2) | 126 (22.3) | | 130 | 32 (15.5) | 98 (22.6) | |
| Female | 547 | 108 (90.8) | 439 (77.7) | | 510 | 175 (84.5) | 335 (77.4) | |
| **Marital status** | | | | **0.000** | | | | **0.000** |
| Single | 422 | 52 (43.7) | 370 (65.5) | | 394 | 102 (49.3) | 292 (67.4) | |
| Married | 262 | 67 (56.3) | 195 (34.5) | | 246 | 105 (50.7) | 141 (32.6) | |
| **Age (years)** | | | | **0.000** | | | | **0.000** |
| ≤35 | 387 | 45 (38.1) | 342 (61.3) | | 355 | 69 (34.2) | 286 (66.4) | |
| 36–60 | 213 | 58 (49.2) | 155 (27.8) | | 205 | 89 (43.1) | 116 (26.9) | |
| >60 | 76 | 15 (12.7) | 61 (10.9) | | 75 | 46 (22.8) | 29 (6.9) | |
| **Education** | | | | 0.811 | | | | 0.065 |
| No college education | 558 | 98 (82.4) | 460 (81.4) | | 119 | 47 (22.7) | 72 (16.6) | |
| College education | 126 | 21 (17.6) | 105 (18.6) | | 521 | 160 (77.3) | 361 (83.4) | |
| **Food security** | | | | 0.957 | | | | 0.486 |
| Food secure | 205 | 37 (33.0) | 168 (32.2) | | 208 | 67 (32.4) | 141 (32.6) | |
| At risk | 198 | 32 (28.6) | 198 (28.6) | | 232 | 81 (39.1) | 151 (34.9) | |
| Experiencing hunger | 199 | 43 (38.4) | 156 (29.9) | | 200 | 59 (28.5) | 141 (32.6) | |
| **Household monthly income (ZAR)** | | | | 0.382 | | | | 0.370 |
| Less than 3000 | 508 | 85 (71.4) | 423 (75.3) | | 478 | 150 (72.8) | 328 (76.1) | |
| More than 3000 | 173 | 34 (28.6) | 139 (24.7) | | 159 | 206 (27.2) | 103 (23.9) | |
| **Number of meals eaten per day** | | | | 0.082 | | | | 0.642 |
| One | 30 | 2 (1.7) | 28 (5.0) | | 30 | 10 (4.9) | 20 (4.6) | |
| Two | 321 | 50 (42.4) | 271 (48.0) | | 299 | 89 (43.2) | 210 (48.5) | |
| Three | 302 | 63 (53.4) | 239 (42.4) | | 283 | 97 (47.1) | 186 (43.0) | |
| Four and more | 29 | 3 (2.5) | 26 (4.6) | | 27 | 17 (3.9) | 10 (4.9) | |
| **Eating of fruit and vegetables** | | | | 0.373 | | | | **0.047** |
| Yes | 634 | 108 (90.8) | 526 (93.1) | | 591 | 185 (89.4) | 406 (93.8) | |
| No | 50 | 11 (9.2) | 39 (6.9) | | 49 | 22 (10.6) | 27 (6.2) | |
| **Eating partners at home** | | | | 0.451 | | | | 0.088 |
| Alone | 541 | 93 (78.2) | 448 (79.3) | | 507 | 174 (84.1) | 333 (76.9) | |
| Parent | 35 | 6 (5.0) | 29 (5.1) | | 34 | 9 (4.3) | 25 (5.8) | |
| Sibling | 60 | 8 (6.7) | 52 (9.2) | | 52 | 16 (7.7) | 36 (8.3) | |
| Combination | 48 | 12 (10.1) | 36 (6.4) | | 47 | 8 (3.9) | 39 (9.0) | |
| **Walking to work** | | | | 0.858 | | | | 0.086 |
| Yes | 74 | 12 (11.0) | 62 (87.8) | | 74 | 20 (10.0) | 54 (12.8) | |
| No | 543 | 97 (89.0) | 446 (87.8) | | 549 | 180 (90.0) | 369 (87.2) | |
| **Dietary diversity score** | | | | **0.025** | | | | 0.767 |
| Low | 68 | 8 (22.2) | 60 (37.7) | | 70 | 20 (36.4) | 50 (34.7) | |
| Medium | 106 | 20 (55.6) | 86 (54.1) | | 108 | 28 (50.9) | 80 (55.6) | |
| High | 21 | 8 (22.2) | 13 (8.2) | | 21 | 7 (12.7) | 14 (9.7) | |

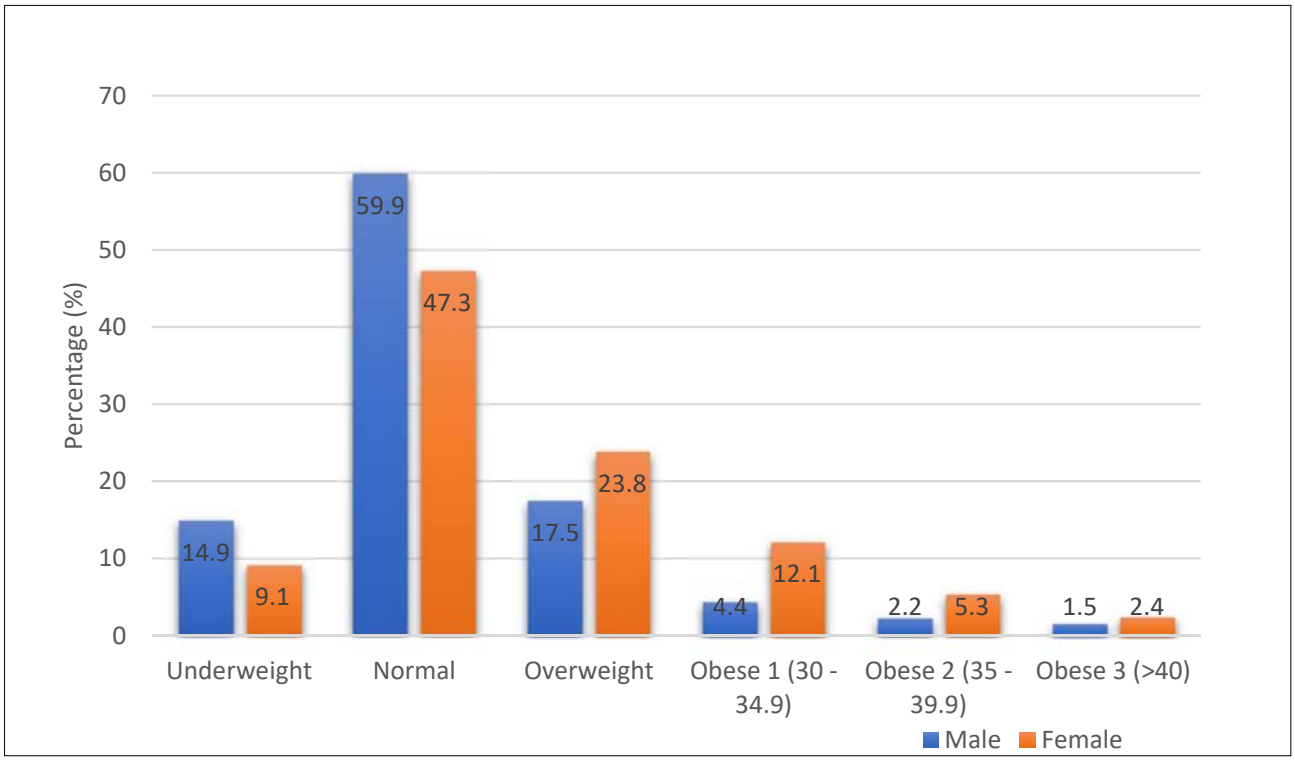



**Figure 1:** Distribution of body mass index by gender ($n$=640).

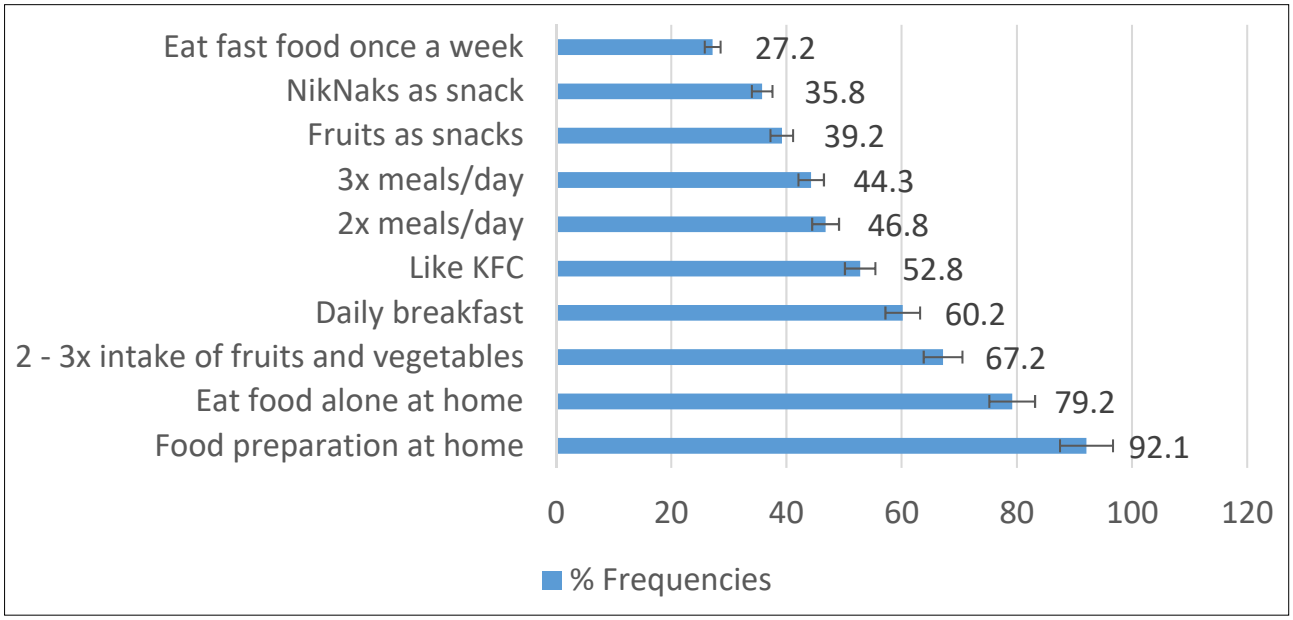



**Figure 2:** Dietary patterns of participants.

**Table 2:** Food groups consumed by >50% of households by dietary diversity category

| Low dietary diversity: ≤3 food groups | Medium dietary diversity: 4–5 food groups | High dietary diversity: ≥6 food groups |
|---|---|---|
| **26.6% of households** | **49.9% of households** | **24.5% of households** |
| Cereal | Cereal | Cereal |
| | Dark green leafy vegetables | Vitamin A rich vegetables and tubers |
| | | Other vegetables |
| | | Oils and fats |

**Table 3:** Association between participant's sociodemographic parameters and dietary diversity score of ≤4 food groups ($n$=640)

| Measures | Crude odds ratio (95% CI) | *p*-value | Adjusted risk ratio (95% CI) | *p*-value |
|---|---|---|---|---|
| **Gender** | | | | |
| Male | 2.00 (0.82–4.57) | 0.127 | 1.92 (0.76–4.87) | 0.167 |
| Female | 1.00 | | 1.00 | |
| **Age (years)** | | | | |
| ≤35 | 0.21 (0.05–0.95) | 0.043 | 0.13 (0.03–0.63) | **0.012** |
| 36–60 | 0.17 (0.04–0.79) | 0.024 | 0.11 (0.02–0.58) | **0.009** |
| >60 | 1.00 | | 1.00 | |
| **Marital status** | | | | |
| Single | 1.12 (0.61–2.06) | 0.717 | 1.03 (0.52–2.04) | 0.934 |
| Married | 1.00 | | 1.00 | |
| **Household monthly income (ZAR)** | | | | |
| Less than 3000 | 0.96 (0.49–1.89) | 0.911 | 2.44 (1.12–5.34) | **0.025** |
| More than 3000 | 1.00 | | 1.00 | |
| **Education** | | | | |
| No college education | 0.89 (0.38–2.06) | 0.785 | 1.27 (0.50–3.20) | 3.204 |
| College education | 1.00 | | 1.00 | |

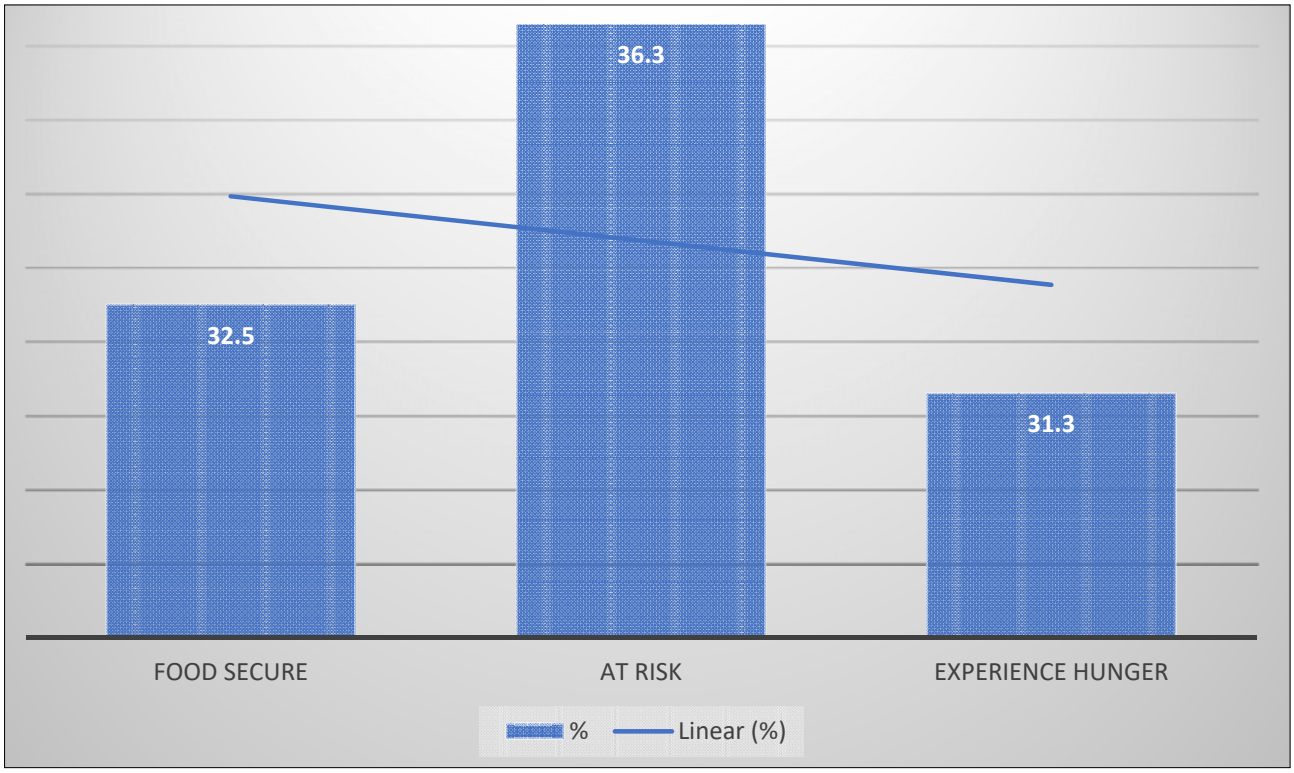



**Figure 3:** Household food security status.

### *Associations between participants' health profile, dietary patterns and food security*

The association between the participants' characteristics, obesity and hypertension are given in Table 1. Being a woman was significantly associated with obesity ($p=0.000$), with 90.8% of obese participants being women, and 9.2% men. In addition, being married and in the age group of 36–60 years were also found to be significantly associated with obesity. Furthermore, age ($p=0.000$), marital status ($p=0.042$), and a healthy eating lifestyle such as daily eating of fruit and vegetables were significantly associated with hypertension status in the current study (Table 1). There was a significant weak positive correlation between BMI and systolic blood pressure severity ($r=0.13$, $p<0.001$). The participants with higher BMI were found to have significantly higher levels of systolic blood pressure as compared to those with lower BMI. We also found a significant association between DDS and obesity ($p=0.025$) and no significant influence of food insecurity on obesity and hypertension status. Walking to work was not associated with obesity.

## Discussion

The main aim of the study was to assess the influence of food insecurity and other predictors on obesity and hypertension status among adults in the Limpopo Province of South Africa. Over one-third of the participants (35.2%) were obese or overweight, with an average BMI of 25.1 kg/m$^2$. Being a woman, older and married were significantly associated with obesity in the current study. The possible reasons could be that women are responsible for food distribution in the household, while older and married persons could have more access to food. In another South African study conducted to determine the risk factors of obesity among the population, it was also found that the female gender and being married were associated with a high BMI.[19] Furthermore, current findings reveal a significantly weak positive correlation between BMI and blood pressure severity ($p=0.001$). This implies that obesity influences hypertension status in the study area. When persons are obese or overweight, the heart has to work harder to pump blood through the body, exerting strain on arteries, resulting in resistance to the flow of blood, causing blood pressure to rise. Previous cohort studies conducted in four sub-Saharan African countries indicated that BMI was one of the primary factors found to be associated with hypertension.[20] Similarly, another study conducted by the United States Health and Nutrition Examination Surveys reported that a positive correlation between BMI and hypertension with BMI$\geq$30 kg/m$^2$ was more prevalent.[21] Thus, increased BMI might lead to increases in blood pressure and cardiovascular problems, proposing the need for measures of better prevention of this increase.[22]

The overall prevalence of hypertension in this study was 32.3% of adults, using the 140/90 mmHg diagnostic threshold of which only 19.3% of participants were aware and treated for the condition. Socio-demographic variables such as gender, age and marital status, and healthy eating behaviour such as the daily intake of fruit and vegetables were found to influence the prevalence of hypertension among the study sample. The current findings are in line with previous studies conducted in countries with high prevalence of hypertension, including the WHO SAGE (wave 1 and wave 2) data which reported age as the major determinant of hypertension among South Africans, with 38% and 42% of participants being unaware of their conditions in wave 1 and wave 2, respectively.[23,24] Further analysis revealed that eating fruit and vegetables was a preventative factor in hypertension, and adds to the increasing body of research supporting the relationship between fruit and vegetable intake and hypertension.[25]

The mean score of the DDS of the current study was 3.99 and the participants' age and household income were found to be the major determinants after adjusting for other factors. This is similar to the national mean DDS of 4.02.[26] Cereal was the main food group consumed by those with the lowest dietary diversity. The present findings show a better DDS when compared to a South African study conducted on the low-income elderly in Sharpeville that delivered a mean DDS of 3.41. The mean DDS delivered in the current study seems lower when compared to studies conducted in other developing countries like the Philippines (mean DDS of 4.91)[27] and Burkina Faso (mean of DDS 4.6).[28] We found a significant inverse correlation between food security and dietary diversity, as did a study conducted in Sekhukhune in Limpopo, and other studies.[29-31] Food insecure households have a limited ability to purchase healthier food options which are often more costly.

The prevalence of food insecurity in the current study was found to be higher than the national rate. Our findings reveal that 36.3% of participants were at risk and 31.3% experienced hunger as compared to the national rate of 28.3% and 26.0%, respectively, in 2012, according to the South African National Health and Nutrition Examination Survey statistics.[18] These findings were slightly lower when compared to another study conducted in one district of the Limpopo Province.

In addition, food insecurity was not associated with hypertension in this study. This might be because the sample was homogeneous. These findings are consistent with an earlier study by Seligman et al.[14] who reported that food insecurity appears to be more strongly associated with diabetes than with hypertension, specifically at the most severe levels of food insecurity.[14] No significant association was found between food security and obesity in the current study. Other studies, however, have also shown no significant association between food security and obesity among the participants.[32,33] These findings contradict results of a study conducted by Gipson-Jones et al.[34] who reported that food insecure parents were more likely to have overweight/obese children.[34] This might be due to the fact that food insecure families have limited ability to purchase healthier food options.

This study was limited because causality cannot be examined because of the cross-sectional study design. Therefore, we cannot attribute cause and effects. Secondly, data were collected using a convenient sample of adults in the community and, therefore, results may not be generalised to the larger population. We did not collect data on smoking and alcohol use, which are main determinants of health risk. Thirdly, data were self-reported, thus social desirability may have affected the extent to which participants reported food-related behaviours in their household. Finally, a single blood pressure measurement is generally not sufficient for a diagnosis of hypertension.

## Conclusion and recommendations

Food insecurity, obesity and hypertension rates remain high among adults in the Limpopo Province of South Africa. Current findings suggest that one in three households are food insecure. Being a woman of advanced age and being single was found to significantly influence obesity and hypertension rates among adults in the study area. Intake of fruit and vegetables was found to positively influence hypertension status. Therefore, educational and nutritional interventions should be designed and targeted towards promoting fruit and vegetable intake in the community. Screening campaigns should be planned to screen asymptomatic individuals for overweight/obesity, elevated blood pressure and hypertension disorders.

## Acknowledgements

The project was funded by the University of Venda.

## Competing interests

We have no competing interests to declare.

## Authors' contributions

X.G.M. conceptualised the study. H.V.M., N.S.M., T.B.A., T.C.M., L.F.M., M.M. and X.G.M. collected the data, analysed the data, wrote the draft and approved final versions.

## References

1. Yew Wong Chin V, Azlan Mohd Noor N. Sociocultural determinants of health and illness: A theoretical inquiry. Geografia Malays J Soc Space. 2014;10(1):49–59. http://ejournal.ukm.my/gmjss/article/view/18364
2. Kgaphola M, Viljoen A. Food habits of rural Swazi households: 1939-1999. Part 2: Social structural and ideological influences on Swazi food habits. J Consum Sci. 2010;32(1):16–25. https://doi.org/10.4314/jfecs.v32i1.52856

3. López AJ. Postcolonial whiteness: A critical reader on race and empire. Albany, NY: State University of New York Press; 2005.

4. Gittelsohn J, Vastine A. Sociocultural and household factors impacting on the selection, allocation and consumption of animal source foods: Current knowledge and application. J Nutr. 2003;133(11):4036S–4041S. https://academic.oup.com/jn/article-abstract/133/11/4036S/4818067

5. Popkin BM. The nutrition transition in the developing world. Dev Policy Rev. 2003;21(5–6):581–597. https://doi.org/10.1111/j.1467-8659.2003.00225.x

6. Harris J, Chisanga B, Drimie S, Kennedy G. Nutrition transition in Zambia: Changing food supply, food prices, household consumption, diet and nutrition outcomes. J Food Secur. 2019;11(2):371–387. https://doi.org/10.1007/s12571-019-00903-4

7. Bärnighausen T, Welz T, Hosegood V, Bätzing-Feigenbaum J, Tanser F, Herbst K, et al. Hiding in the shadows of the HIV epidemic: Obesity and hypertension in a rural population with very high HIV prevalence in South Africa. J Hum Hypertens. 2008;22(3):236. https://www.nature.com/articles/1002308

8. Statistic South Africa. Mortality and Causes of Death in South Africa, 2013: Findings from Death Notification. Pretoria; 2014.

9. Whitaker R, Phillips S, Orzol S. Food insecurity and the risks of depression and anxiety in mothers and behavior problems in their preschool-aged children. Pediatrics. 2006;118:e859–868. https://doi.org/10.1542/peds.2006-0239

10. Slack KS, Yoo J. Food hardship and child behavior problems among low-income children. Soc Serv Rev. 2005;79(3):511–536. https://doi.org/10.1086/430894

11. Kursmark M, Weitzman M. Recent findings concerning childhood food insecurity. Current Opinion in Clinical Nutrition Metabolic care. 2009;12:310–316. https://journals.lww.com/co-clinicalnutrition/Fulltext/2009/05000/Enteral_feeding_and_gut_atrophy.16.aspx?casa_token=0eUgbuMaWGcAAAAA:9qk1Da_rPyMNp-kAkPS0HxD5bqaQ55TM3XsXNLhkXQCt_7cy4BbFLlnwnnuNPdoV1FrpSWFm9p5t02ce_XxOcOkvIuWKvCRDvA.

12. Skalicky A, Meyers AF, Adams WG, Yang Z, Cook JT, Frank DA. Child food insecurity and iron deficiency anemia in low-income infants and toddlers in the United States. Matern Child Health J. 2006;10(2):177–185. https://doi.org/10.1007/s10995-005-0036-0

13. Rose-Jacobs R, Black MM, Casey PH, Cook JT, Cutts DB, Chilton M, et al. Household food insecurity: Associations with at-risk infant and toddler development. Pediatrics. 2008;121:65–72. https://doi.org/10.1542/peds.2006-3717

14. Seligman H, Laraia B, Kushel M. Food insecurity is associated with chronic disease among low-income NHANES participants. J Nutr. 2009;140(2):304–310. https://doi.org/10.3945/jn.109.112573

15. Dinour L, Bergen D, Yeh M. The food insecurity–obesity paradox: A review of the literature and the role food stamps may play. J Am Diet Assoc. 2007;107(11):1952–1961. https://doi.org/10.1016/j.jada.2007.08.006

16. Kushel MB, Gupta R, Gee L, Haas JS. Housing instability and food insecurity as barriers to health care among low-income Americans. J Gen Intern Med. 2006;21(1):71–77. https://doi.org/10.1111/j.1525-1497.2005.00278.x

17. Labadarios D, Steyn NP, Maunder E, MacIntyre, U, Gericke G, Swart R, et al. The National Food Consumption Survey (NFCS): South Africa. Public Health Nutr. 1999;8(5):533–543. https://doi.org/10.1079/PHN2005816

18. Shisana O, Labadorias D, Rehle T, Simbayi L, Zuma K, Dhansay A. South Africa National Health and Nutrition Examination Survey (SANHANES-1). Cape Town: HSRC Press; 2013.

19. Malhotra R, Hoyo C, Østbye T, Hughes G, Schwartz D, Tsolekile L, et al. Determinants of obesity in an urban township of South Africa. S Afr J Clin Nutr. 2008;21(4):315–320. https://doi.org/10.1080/16070658.2008.11734173

20. Baiyegunhi LJS, Makwangudze KE. Home gardening and food security status of HIV/AIDS affected households in Mpophomeni, KwaZulu-Natal Province, South Africa. Hum Ecol. 2013;44(1):1–8. https://doi.org/10.1080/09709274.2013.11906637

21. Wang Y, Wang QJ. The prevalence of prehypertension and hypertension among US adults according to the new joint national committee guidelines: New challenges of the old problem. Arch Intern Med. 2004;164(19):2126–2134. https://doi.org/10.1001/archinte.164.19.2126

22. Linderman GC, Lu J, Lu Y, Sun X, Xu W, Nasir K, et al. Association of body mass index with blood pressure among 1.7 million Chinese adults. JAMA Netw Open. 2018;1(4), e181271. https://doi.org/10.1001/jamanetworkopen.2018.1271

23. Lloyd-Sherlock P, Beard J, Minicuci N, Ebrahim S, Chatterji S. Hypertension among older adults in lowand middle-income countries: Prevalence, awareness and control. Int J Epidemiol. 2014;43(1):116–128. https://doi.org/10.1093/ije/dyt215

24. Ware LJ, Chidumwa G, Charlton K, Schutte AE, Kowal P. Predictors of hypertension awareness, treatment and control in South Africa: Results from the WHO-SAGE population survey (Wave 2). J Hum Hypertens. 2019;33(2):157–166. https://doi.org/10.1038/s41371-018-0125-3

25. Wang L, Manson JE, Gaziano JM, Buring JE, Sesso HD. Fruit and vegetable intake and the risk of hypertension in middle-aged and older women. Am J Hypertens. 2012;25(2):180–189. https://doi.org/10.1038/ajh.2011.186

26. Labadarios D, Steyn N, Nel J. How diverse is the diet of adult South Africans. Nutr J. 2011;10(33):1–11.

27. Kennedy G, Pedro M, Seghieri C, Nantel G, Brouwer ID. Dietary diversity score is a useful indicator of micronutrient intake in non breast-feeding Filipino children. J Nutr. 2007;137:472–477.

28. Kennedy G. Evaluation of dietary diversity scores for assessment of micronutrient intake and food security in developing countries [PhD thesis]. Wageningen: Wageningen University; 2009.

29. Faber M, Schwabe C, Drimie S. Dietary diversity in relation to other household food security indicators. Int J Food Saf Nutr Publ Health. 2008;1(2):157–171.

30. Thornton AJ. Dietary diversity and food security in South Africa: An application using NIDS Wave 1 [masters thesis]. Cape Town: University of Cape Town; 2016. https://open.uct.ac.za/handle/11427/20617

31. Psaki S, Bhutta ZA, Ahmed T, Ahmed S, Bessong P, Islam M, et al. Household food access and child malnutrition: Results from the eight-country MAL-ED study. Popul Health Metr. 2012;10, Art. #24. https://doi.org/10.1186/1478-7954-10-24

32. Wake M, Salmon L, Waters E, Wright M, Hesketh K. Parent-reported health status of overweight and obese Australian primary school children: A cross-sectional population survey. Int J Obes Relat Metab Disord. 2002;26(5):717.

33. Young-Hyman D, Herman LJ, Scott DL, Schlundt DG. Care giver perception of children's obesity-related health risk: A study of African American families. Obes Res. 2000;8(3):241–248. https://doi.org/10.1038/oby.2000.28

34. Gipson-Jones T, O'Neal L, Sheats J, Thorpe JR, Beech B, Bruce MA. Food security status and overweight/obesity among 2-to 5-year-old boys and girls in a community-based clinic. Fam Community Health. 2019;42(2):117–122. https://doi.org/10.1097/FCH.0000000000000218

**AUTHORS:**
Njongenhle M.B. Nyoni[1,2]
Stefan Grab[1]
Emma Archer[3]
Johan Malherbe[4,5]

**AFFILIATIONS:**
[1]School of Geography, Archaeology and Environmental Studies, University of the Witwatersrand, Johannesburg, South Africa
[2]Food, Agriculture and Natural Resources Policy Analysis Network (FANRPAN), Pretoria, South Africa
[3]Department of Geography, Geoinformatics and Meteorology, University of Pretoria, Pretoria, South Africa
[4]Soil, Climate and Water – Agricultural Research Council, Pretoria, South Africa
[5]Smart Places, Council for Scientific and Industrial Research (CSIR), Johannesburg, South Africa

**CORRESPONDENCE TO:**
Njongenhle Nyoni

**EMAIL:**
nmbnyoni@gmail.com

**DATES:**
**Received:** 23 Jan. 2020
**Revised:** 10 June 2021
**Accepted:** 12 July 2021
**Published:** 29 Nov. 2021

**HOW TO CITE:**
Nyoni NMB, Grab S, Archer E, Malherbe J. Temperature and relative humidity trends in the northernmost region of South Africa, 1950–2016. S Afr J Sci. 2021;117(11/12), Art. #7852. https://doi.org/10.17159/sajs.2021/7852

**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:**
agriculture, climate change, temperature, humidity, northernmost South Africa

**FUNDING:**
Open Society Foundation

# Temperature and relative humidity trends in the northernmost region of South Africa, 1950–2016

The northernmost Limpopo Province is located in one of the warmest regions of South Africa, where the agricultural sector is prone to heat stress. The aim of this study was to explore air temperature and relative humidity trends for the region, which have implications for agricultural adaptation and management (amongst other sectors). In particular, we investigated seasonal, annual and decadal scale air temperature and relative humidity changes for the period 1950–2016. Positive temperature trends were recorded for this period, averaging +0.02 °C/year, with the strongest changes observed in mean maximum summer temperatures (+0.03 °C/year). Interannual temperature variability also increased over time, especially for the period 2010–2016, which presents probability densities of <50% for minimum temperatures. Positive relative humidity trends (+0.06%/year) were also recorded for the period 1980–2016, but proved to be the least predictable weather parameter, with probability densities of <0.5% across seasons for the study period. Considering the substantial interannual variability in temperature and relative humidity, there is clear increased risk for the agricultural sector, particularly for small-scale farmers who generally have limited capacity to adapt. Climate science focusing on the southern African region should continue to establish the impact of climate change and variability on specific small-scale farming systems and enterprises, with recommendations for strategic adaptation based on up-to-date evidence.

**Significance:**

- Heat indices have increased, and variability in temperature and relative humidity has substantially increased over recent decades.
- Changes in air temperature and relative humidity have direct and/or indirect negative effects on sectors such as agriculture, leading to reduced productivity.
- The small-scale farming sector, which contributes significantly to national food security in developing countries, is the production system most exposed and vulnerable to observed changes/extremes in temperature and relative humidity.
- There is an urgent need to build capacity of small-scale farmers for appropriate adaptation to observed changes in climate based on up-to-date evidence.

## Introduction

Globally, mean air temperatures have increased by at least 0.65 °C over the past century, while extreme rainfall events have also increased in frequency.[1-3] It is widely projected that as the earth becomes warmer, climate and weather variability will increase.[4-7] Projected temperatures for southern Africa indicate expected near-future increases at twice the global average.[8] Such changes in climate present a significant threat to food security, water resources, health and infrastructure.[2] Low-income countries are likely to be the most vulnerable to extreme climatic events due to their poor adaptive capacity in terms of finances, resources, infrastructure and expertise.[9] In particular, the small-scale farming sector, which contributes significantly to national food security in developing countries, is an exposed and vulnerable production system.[10] The situation is compounded in cases where there are limited reliable historical climate data available to establish climatic trends – a concern that holds true for most African countries.

In order to develop a sustainable adaptive capacity for target regions, and address current and future challenges presented by climate variability and/or change, it is critical to evaluate historical climate trends.[10] Process-based regional and global climate models generally have the ability to replicate large-scale climate features. However, their coarseness in spatial resolution poses challenges in providing usable information at local scales.[11,12] Region-specific historically observed climatic records, where possible, have the potential to assist ongoing climate modelling and projections, and to establish the likely impact of climate change on agricultural productivity and ultimately food security.

Several studies on historical weather and climate change have been focused at the countrywide level in South Africa[12-14], while others have been more region specific, with finer spatial detail[14-18]. For example, the study by Gbetibouo[15] on the Limpopo Basin area in northernmost South Africa provides some information on climate trends in the region. It should be noted that the study mainly focused on farmers' perceptions and adaptation to climate change and variability, rather than a detailed analysis on climate trends. A further study focused on 30 Limpopo Province catchments, analysing temperature trends for the period 1950–1999[16], and established increases in annual and seasonal temperature. However, this study did not include relative humidity, and hence was unable to establish heat indices. In the context of agricultural and human/animal thermal comfort levels, a combination of these two weather parameters (i.e. temperature and relative humidity) is important for establishing

heat indices – effectively a measure of how hot it feels. Understanding of heat indices is particularly important for most resource-limited rural communities in Limpopo Province and other rural sectors of southern Africa involved in smallholder farming enterprises, many of which may already be adversely affected by extreme weather events. Evidence shows that increases in heat index trends will have far-reaching impacts on agricultural production and productivity, especially in poor communities with constrained adaptive capacity.[19,20] We thus aimed to provide an improved understanding of recent temperature and relative humidity trends in the northernmost region of the Limpopo Province, which has implications for determining heat stress. The particular focus here was on analysing seasonal and annual temperature and relative humidity trends for the period 1950–2016, and hence the computation of heat indices.

## Materials and methods

### *Study area*

This study focused on the northern border region of the Limpopo Province, South Africa (Figure 1). The region is located in the Vhembe District Municipality, where a large portion of its rural population resides. This semi-arid region is located in one of the warmest parts of South Africa where temperatures regularly exceed 40 °C during the warm season.[21] Mean daily maximum temperatures during summer also exceed 30 °C, and precipitation averages only ~250 mm p.a., most of which falls during summer.[21] Such climatic conditions (particularly high temperatures) regularly cause heat stress, especially in the livestock sector, thereby reducing potential productivity.[22] The Limpopo Province is one of the economically and financially poorest provinces in the country, with livelihoods in most rural villages depending on subsistence farming and casual employment. Government grants and remittances from off-site relatives contribute significantly to household incomes. Smallholder farmers produce crops and rear some livestock.

### *Data collection*

Daily temperature (1950–2016) and relative humidity (1980–2016) data for two weather stations (Macuville: data period = 1950–2014; Venetia Mine: data period = 2015–2016) in northernmost Limpopo Province were provided by the South African Weather Service. Temperature and humidity values are variable across any given landscape due to topography, land cover and external climatic inputs (e.g. airflow from surrounding regions). The values we present here are thus in *absolute* terms, only relevant to the station localities, and for the broader region serve as a more general (*relative*) indication of conditions, rather than *absolute* conditions for any given site in the province. Information was collated monthly and seasonally. Traditional seasons in South Africa are divided into summer (December – February), autumn (March – May), winter (June – August), and spring (September – November).[4] For ease of analysis, temperature data were further divided into two periods: 1950–1986 and 1980–2016.

Data were assessed to identify missing and incorrectly reported or recorded values (e.g. daily minimum temperature greater than the daily maximum temperature). Missing and/or incorrect values were then replaced using temporal interpolation techniques (hierarchical polynomial regression techniques, in particular) as described by Boissonnade et al.[23] Missing values were replaced through interpolations between observations over time (temporal interpolation). In addition, change points were identified[24] and homogenised using the quantile-matching adjustment method[25]. Using the quantile-matching method, up to 10 years' of data, before or after a change point, were used to produce reliable adjustments.[24,25] For relative humidity, zero drift in sensors was assumed.

### *Data analyses*

Temperature and relative humidity trends were determined using R software packages (lubridate and forecast)[26,27] with seasonal and trend decomposition using the locally weighted smoothing (LOESS) (STL) function. LOESS was used in regression analysis for creating a smooth line through a timeplot, thus demonstrating the relationship between variables and forecast trends. Graphs plotted from STL show four components: (1) the original data (i.e. a set of actual values); (2) a seasonal component calculated using LOESS smoothing; (3) the trend (i.e. increasing or decreasing direction in the data); and (4) the remainder representing the residual.

The total number, amplitude and intensity of heatwave days were calculated for each year. Threshold values for a heatwave were based on the average maximum temperature of the hottest month of a given year plus 5 °*C, as described by* Mbokodo et al.[28] The totals per year were only for heatwave days, where the maximum temperatures exceeded the threshold for three or more successive days, thus guaranteeing that heatwave events with a shorter duration were also detected. It is worth noting, however, that this heatwave computation ignores extremes associated with cooler months and changes in adaptation-related impacts.

Probability density plots, calculated using R software package ggplot2[29], were tested for significance at the 95% confidence level, and used to establish variability of temperature and relative humidity across the seasons and decades under study. Density plots provide a relative likelihood of these random variables falling within a particular range of values. The heat index (incorporating air temperature and relative humidity) was calculated using weathermetrics[30]. Maximum daily air temperatures and minimum relative humidity were used to determine the heat index. Weathermetrics' heat.index creates a numeric vector of heat index values from numeric vectors of air temperature and either relative humidity or dew point temperature. In calculating the heat index in R using the weathermetrics package, the following code was used:

heat.index (t = NA, dp = c(), rh = c(), temperature.metric = 'celsius', output.metric = celsius),

where t is numeric vector of air temperatures; rh is numeric vector of relative humidity (in %); temperature.metric is the character string indicating the temperature metric of air temperature and dew point temperature (possible values are 'celsius'); and output.metric is the character string indicating the metric into which heat index should be calculated (possible values are 'celsius').

Weathermetrics calculations for heat index are based on the National Weather Service Hydrometeorological Prediction Center Web Team Heat Index Calculator.[31] Results from weathermetrics were validated by the RClimdex version 1.9[32] software calculations on warm days (TX90p) and warm spell duration indicator (WSDI). TX90p shows the percentage of days when maximum temperature (TX) is greater than the 90th percentile centred on a 5-day window, while WSDI highlights the annual count of days with at least 6 consecutive days when TX is greater than the 90th percentile.

## Results

### *Temperature trends*

Time series graphs show positive trends in temperature, with averages marked for each month over the 1950–2016 period (see Figures 2–4). Rates of minimum, mean and maximum temperature increases were similar (+0.02 °C/year) for the study period. However, summer maximum temperatures recorded the highest rate of increase (+0.03 °C/year) compared to other seasons and other temperature parameters (i.e. average and minimum). Seasonal and trend decomposition of temperatures show that mean maximum temperatures ranged between 28 °C and 33 °C (see Supplementary figure 1), average temperatures ranged from 21 °C to 25 °C (see Supplementary figure 2), and mean minimum temperatures from 14 °C to 18 °C (see Supplementary figure 3). The absolute highest and lowest recorded temperatures were during the summer of 1992 (45.2 °C) and winter of 1972 (-3.8 °C), respectively (see Table 1).

Strongest positive warming trends are observed for the past three and a half decades (i.e. from 1980–2016). For instance, mean annual maximum temperatures increased at a rate of +0.02 °C/year between 1950 and 1986, but substantially increased to +0.03 °C/year for the period 1980–2016.


**Figure 1:** The study area: Northernmost Limpopo Province, South Africa.

The rate at which mean annual temperatures have increased, has remained similar over time: 0.02 °C/year over the period 1980–2016, compared to 0.02 °C/year for the period 1950–1986. None of the warming trends (i.e. for the periods 1950–2016, 1950–1986 or 1980–2016) were statistically significant at the 95% confidence level.

### *Heatwave trends*

The value for determining the threshold for a heatwave in northernmost Limpopo Province was 38.9 ºC (i.e. 33.9 ºC+5 ºC). All heatwaves occurred during the summer season. The rate of increase in the number of heatwave occurrences per annum is +0.07 days/year for the study period (see Figure 5). The frequency of heatwaves for the period 1950–1986 was ~1.5/year, which then increased substantially to ~3.1/year for the period 1980–2016. Further, the mean duration of heatwaves increased during the last three decades (mean = 5.9 days) over that of the previous decades (mean = 4.6 days).

The annual heatwave amplitude (1950–2016) during periods associated with a heatwave in northernmost Limpopo Province is presented in Figure 6. The lowest (40.1 ºC) heatwave amplitude was recorded in December of 1988, while the highest (45.2 ºC) was in December of 1992. Generally, heatwave amplitude increased significantly ($p=0.04$) over the study period. The intensity of heatwaves (i.e. average temperature with which the heatwave threshold was exceeded per year, taking all heatwaves for that year into account [see Figure 7]) also increased significantly ($p=0.04$) over the period 1950–2016. Table 2 shows heatwave trends in terms of change in heatwave amplitude and intensity of heatwaves.

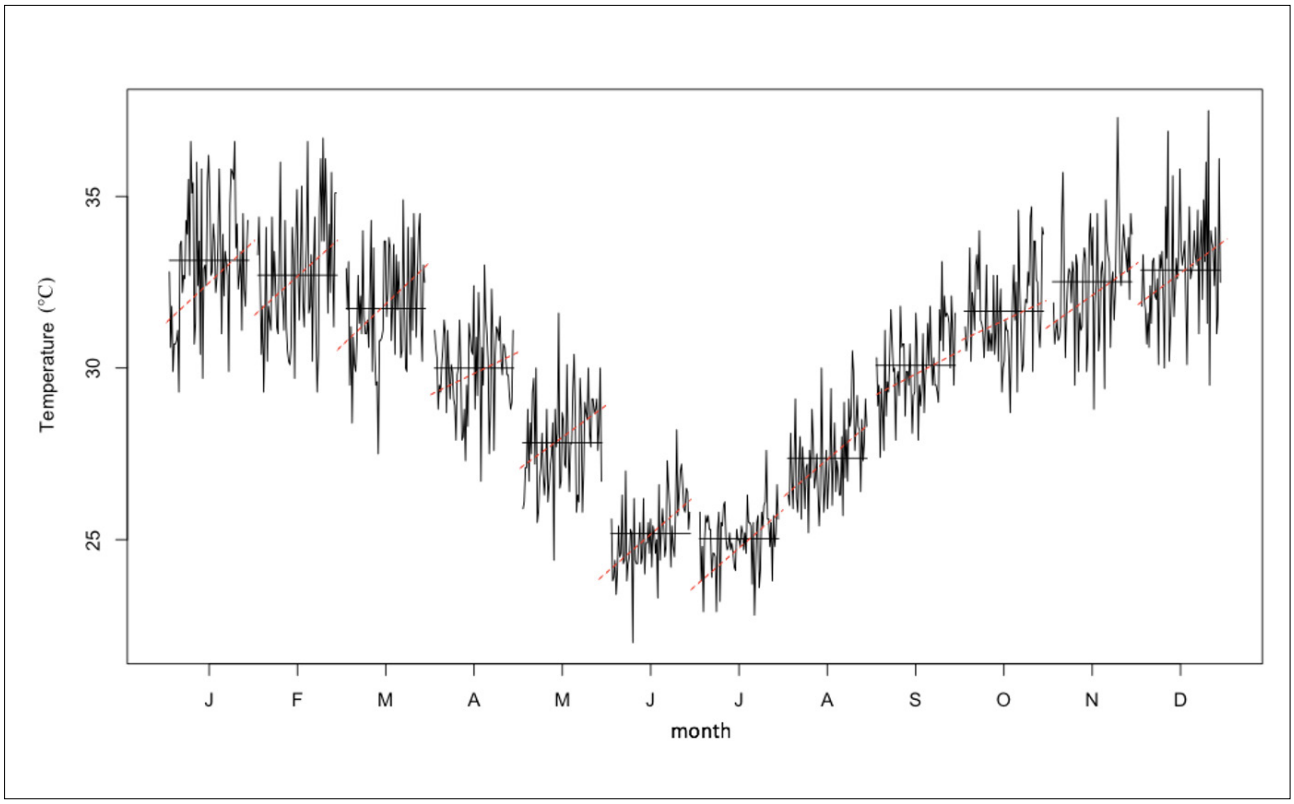



**Figure 2:** Mean monthly maximum temperatures for northernmost Limpopo Province, 1950–2016. Horizontal lines represent monthly mean values and red dotted lines represent trend lines.

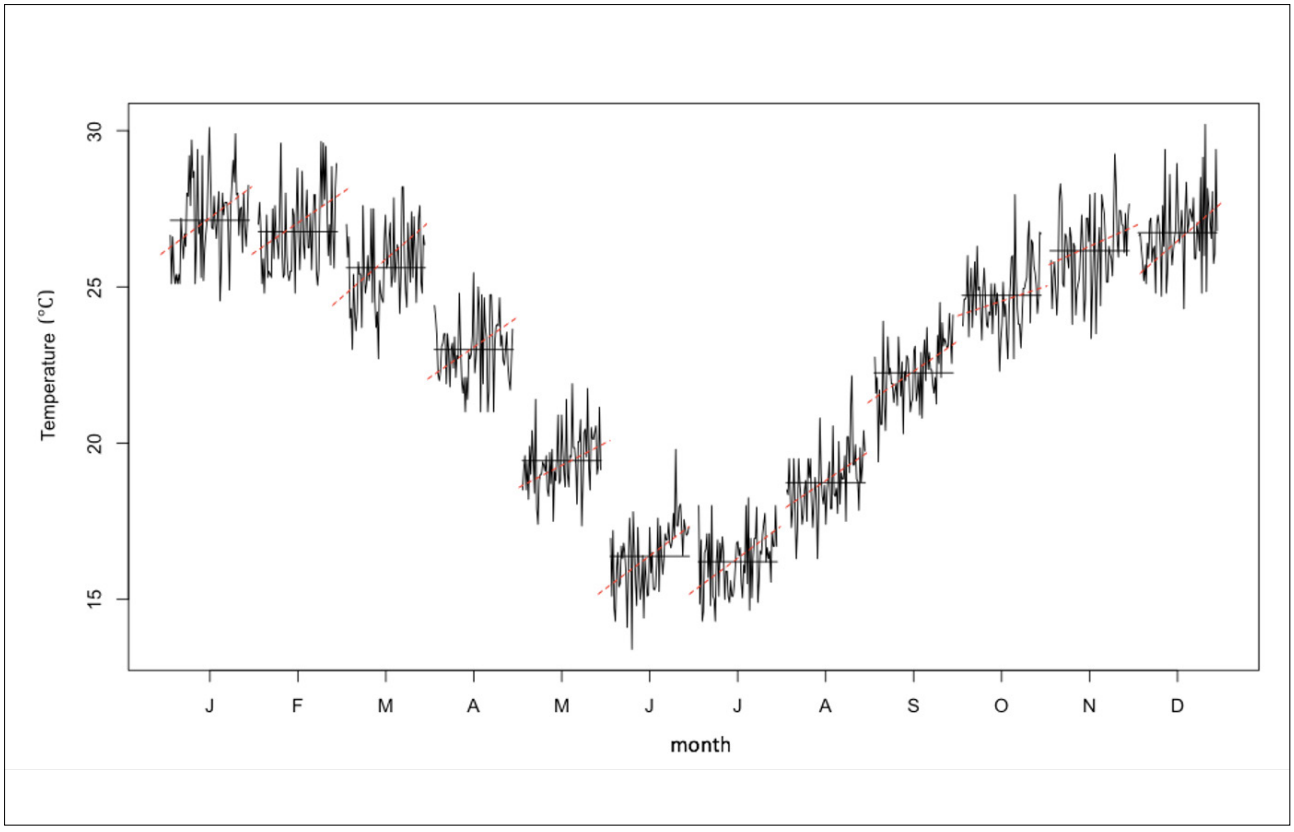



**Figure 3:** Mean monthly temperatures for northernmost Limpopo Province, 1950–2016. Horizontal lines represent monthly mean values and red dotted lines represent trend lines.

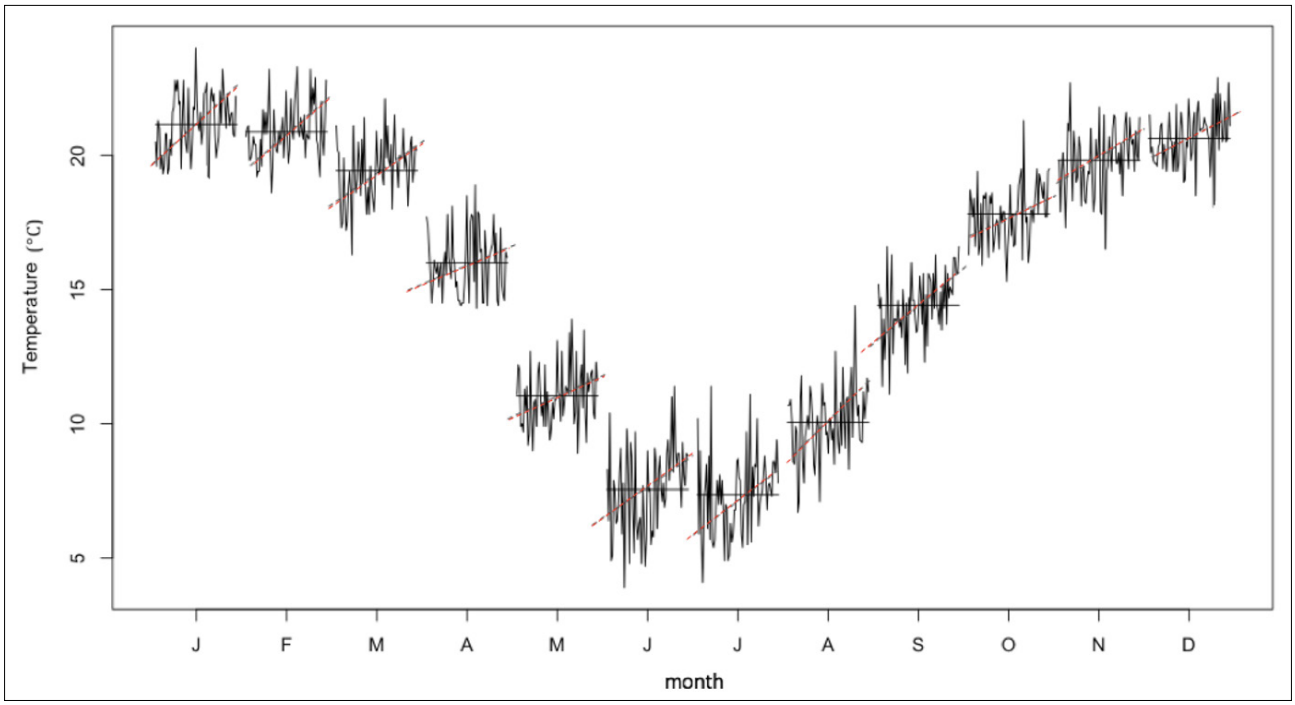



**Figure 4:** Mean monthly minimum temperatures for northernmost Limpopo Province, 1950–2016. Horizontal lines represent monthly mean values and red dotted lines represent trend lines.

**Table 1:** Monthly temperatures for northernmost Limpopo Province, 1950-2016

| | Trends in temperature | Mean temperature range | Maximum | Minimum |
|---|---|---|---|---|
| Maximum temperature | +0.02 °C/year | 28–33 °C | +45.2 °C<br>Summer (1992) | +15.5 °C<br>Winter (2010) |
| Mean temperature | +0.02 °C/year | 21–25 °C | | |
| Maximum temperature | +0.02 °C/year | 14–18 °C | +29.7 °C<br>Summer (2007) | -3.8 °C<br>Winter (1972) |

### *Relative humidity trends*

Results show a positive relative humidity trend (+0.06%/year), albeit not statistically significant ($p=0.3$) for the period 1980–2016, with mean seasonal relative humidity ranging between 10% during winter (July) of 1980, and 83.2% during autumn (April) of 2012. Generally, summers had the highest relative humidity (mean = 46.8%) and winters the lowest (mean = 36.3%) (Table 3). Seasonal relative humidity increased during summer (+0.12%/year) and autumn (+0.10%/year), and decreased during winter (-0.17%/year) and autumn (-0.15%/year) (Figure 8).

### *Variability in temperature and relative humidity*

Weather parameters across all seasons had low probability densities (below 50%), implying high variability (Figure 9). Highest probability densities of maximum temperature were observed for summer over the periods 1980–1989 and 1990–1999, and in autumn over the periods 2000-2009 and 2010-2016. Winter probability densities were lowest over the entire study period, except for the period 2000–2010, when spring recorded the lowest (~12%). The highest (~37%) summer probability densities were recorded for the period 1990–1999, and lowest (15%) over the period 2000–2009. For winter, however, maximum temperature probability densities were consistently at ~15% throughout the study period, indicating high variability.

For average temperatures, the highest and lowest probability densities were recorded for summer and winter respectively, throughout the study period. Highest summer (~52%) probability densities were recorded during the period 1990–1999 and lowest (~30%) during 2000–2009. As with maximum temperatures, winter probability densities were consistently at ~10% for the period 1980–2016. Consistent with probability densities for average temperatures, minimum temperatures were always highest for summer and lowest for winter. Summer probability densities were highest (~41%) over the period 2010-2016, and lowest (~30%) during the period 2000–2019. Again, winter probability densities were consistently lowest at ~15% for the entire study period.

Probability densities of relative humidity across all seasons were always very low (i.e. less than 10%), with considerable inter-annual and decadal variability. The period 2010–2016 presented the highest variations, with probability densities of <0.5% for relative humidity. The highest probability densities (from highest to lowest) were recorded for the period 1980–1989 (autumn), 2000–2009 (spring), 1990–1999 (autumn) and 2010–2016 (spring). Conversely, the lowest probability densities (from lowest to highest) were recorded in winter (2010–2016), summer (2000–2009), summer (1980–1989) and spring (1990–1999), respectively.

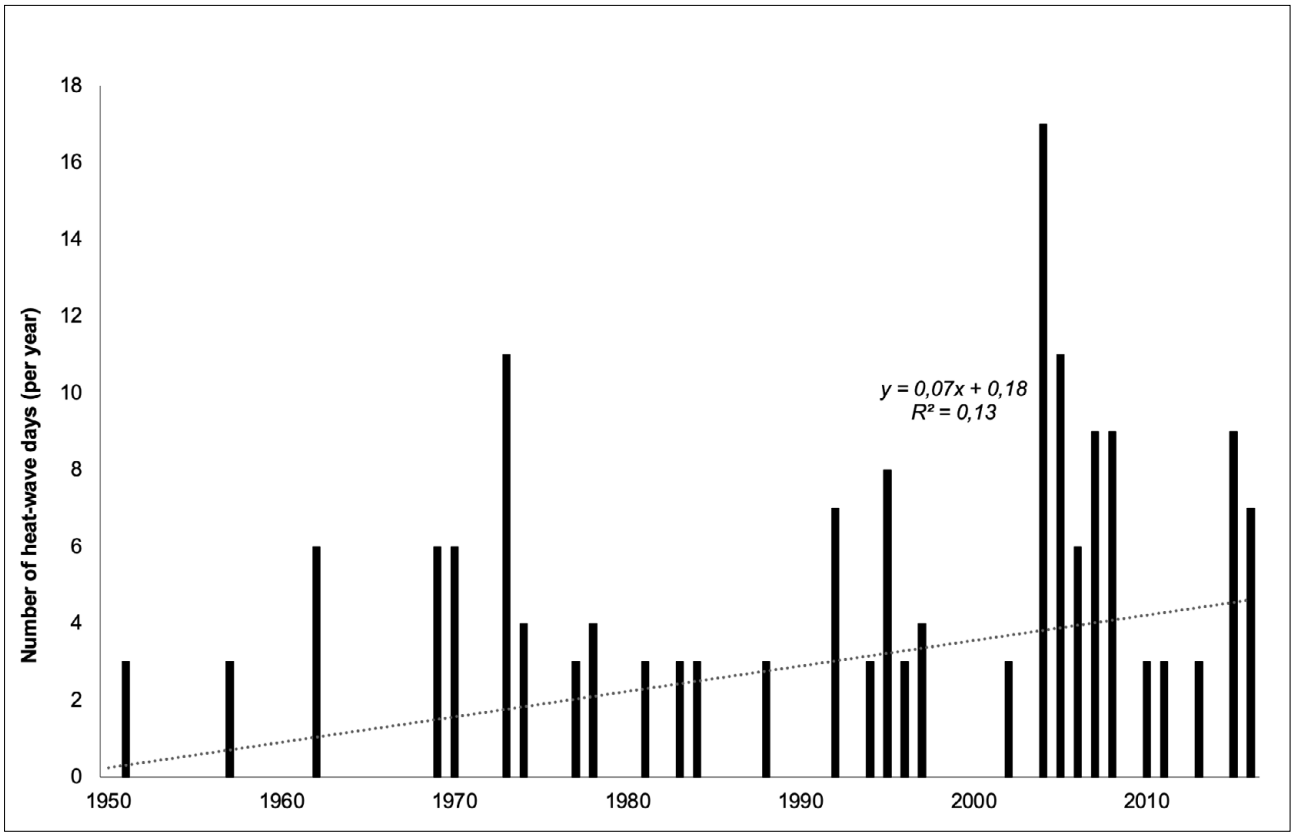



**Figure 5:** Number of heatwave days per annum for northernmost Limpopo Province, based on mean maximum temperature for the hottest month (January). The value for determining the threshold for a heatwave in northernmost Limpopo Province is 38.9 °C (i.e. 33.9 °C + 5 °C).

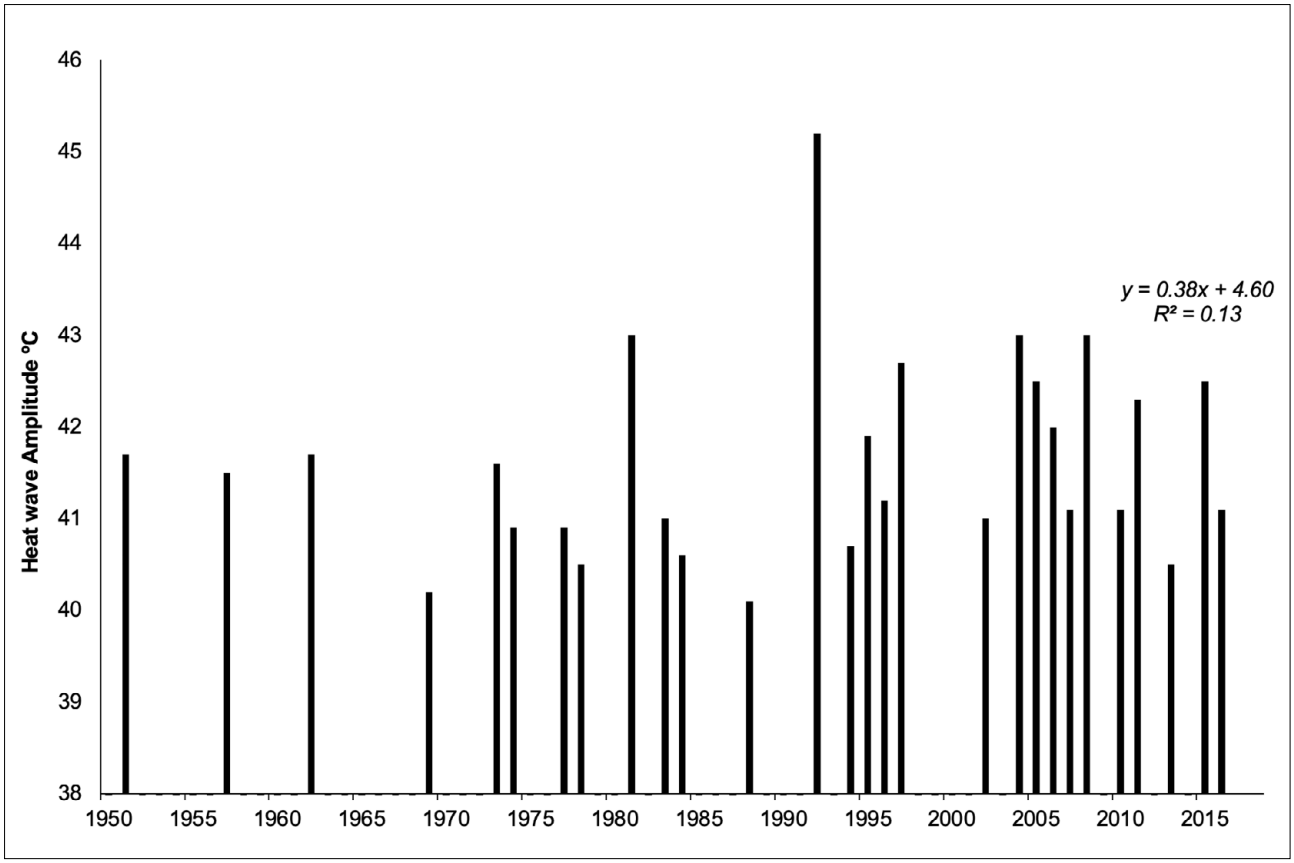



**Figure 6:** Annual heatwave amplitudes (i.e. highest temperature per year during periods associated with a heatwave) for northernmost Limpopo Province (1950–2016).

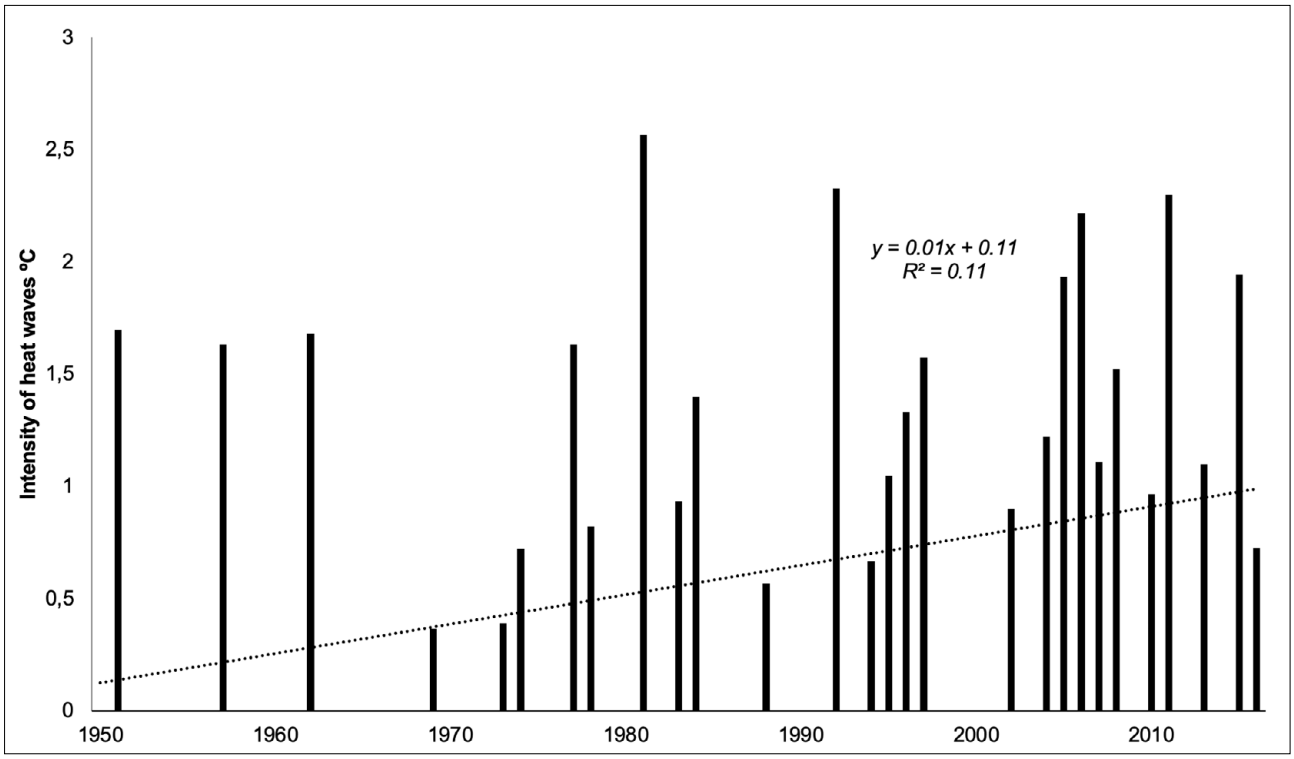



**Figure 7:** Intensity of heatwaves (i.e. average yearly exceedance temperature of heatwave threshold during heatwave days) for northernmost Limpopo Province, plotted against time for the period 1950–2016.

**Table 2:** Heatwave trends for the period 1950–2016 in northernmost Limpopo Province

| | Trend | *p*-value |
|---|---|---|
| Heatwave amplitude | +0.38 °C/year | 0.04 |
| Intensity of heatwaves | +0.01 °C/year | 0.04 |

**Table 3:** Relative humidity seasonal means for northernmost Limpopo Province, 1950–2016

| | Trends | Range | Average | | | |
|---|---|---|---|---|---|---|
| | | | Summer | Autumn | Winter | Spring |
| Relative humidity | +0.06%/year | 10–83.2% | 46.8% | 42.5% | 36.3% | 38.1% |

Of the two variables (temperature and relative humidity), relative humidity was the least predictable weather variable across seasons for the study period. Mean summer temperatures across all decades were significantly ($p=0.03$) more predictable (~50%) than other temperature parameters during summer and other seasons. The only exceptions were predictions for autumn maximum temperatures over the years 2000–2009 and 2010–2016, both having probability densities of ~25%, a value which exceeded the summer percentages.

### *Heat index*

Heat index values (using R Package weathermetrics) were calculated for the period 1980–2016, based on available relative humidity and maximum temperature data. According to seasonal and trend decomposition time series (see Figure 10), the heat index ranged between 24 °C and 35 °C. Highest heat index measurements were observed during the summer of 2007, and the lowest (+24 °C) during the winter of 1996.

Over the study period, positive trends were observed for all seasons (Figure 11) i.e. summer (+0.05 °C/year), autumn (+0.05 °C/year), winter (+0.04 °C/year) and spring (+0.05 °C/year). Overall, there has been a positive heat index trend (+0.05 °C/year). RClimdex calculations validating these results are provided in Supplementary figures 4 (TX90p) and 5 (WSDI).

## Discussion

Our analysis shows that there has been a positive mean annual warming trend (+0.02 °C/year) for northernmost Limpopo Province over the period 1950–2016, which is in line with what other recent studies in the southern African region have reported.[12,14,33,34] Temperature increases have also resulted in increased frequency, amplitude and intensity of heatwaves and a positive heat index trend (+0.03 °C/year). Climate studies from other southern African regions[1,4,28,35-37] have also reported increased frequency of heatwaves and hot days, associated with accelerated global warming. The primary concern with such climatic changes is that they significantly impact sectors such as agriculture and water resources, directly and/or indirectly, reducing productivity.[37,38] Thus, the rapid increase in heatwave frequency and intensity over northernmost South Africa poses major concerns to agriculture in an already water-scarce environment, and even more so, given that evaporation rates and water consumption demands have increased.[8,28] High temperatures, especially when combined with high humidity and low air movement, can exceed species-specific threshold levels (or thermal comfort zones), thereby causing heat stress.[39] It is likely that farming systems in the northern border region of Limpopo Province are already adversely impacted by climate change, particularly so in small-scale subsistence settings where farmers have limited capacity to adapt. This is in line with findings from an earlier study in Limpopo Province, which established that, although farmers were able to recognise climate change, only a few employed some form of adaptation strategy.[15]

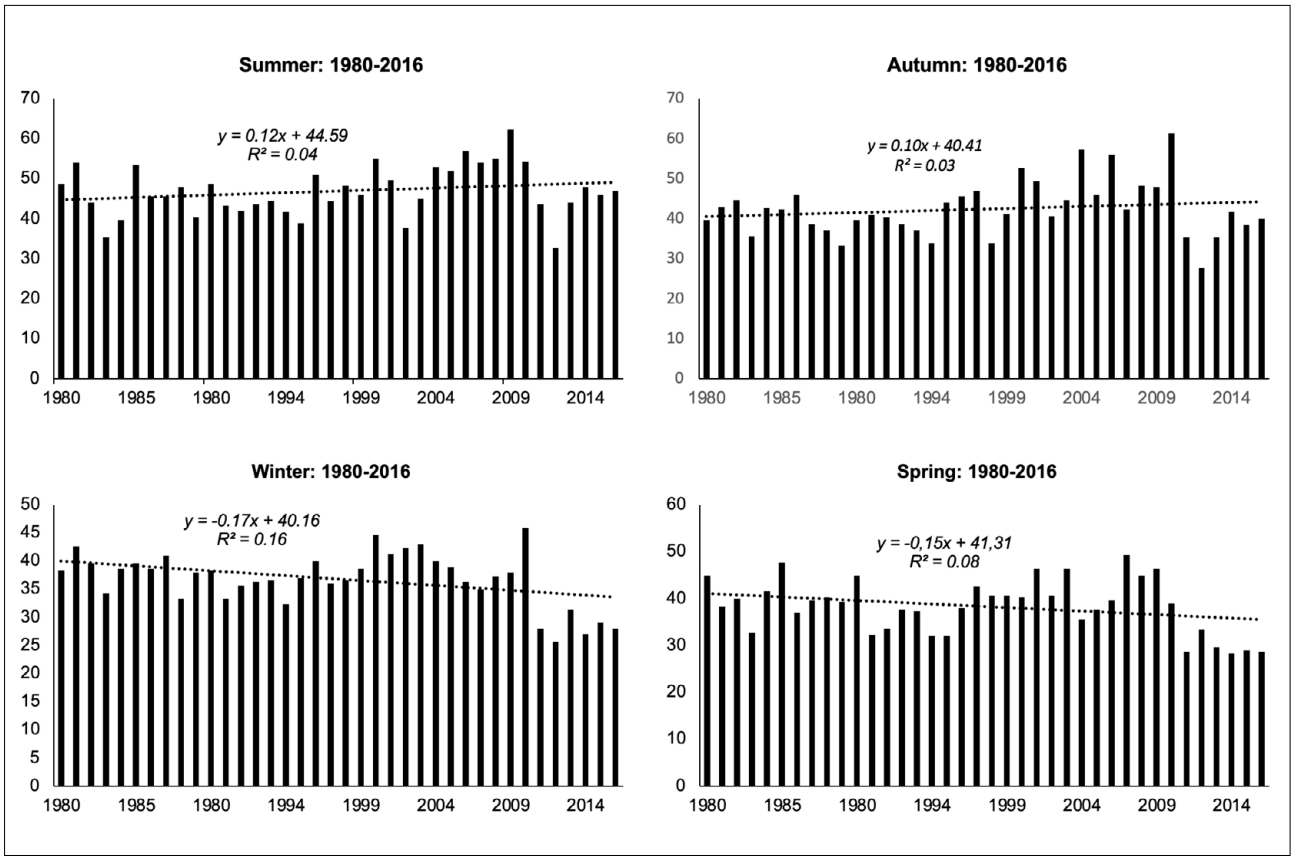



**Figure 8:** Mean seasonal relative humidity trends for northernmost Limpopo Province, 1980–2016.

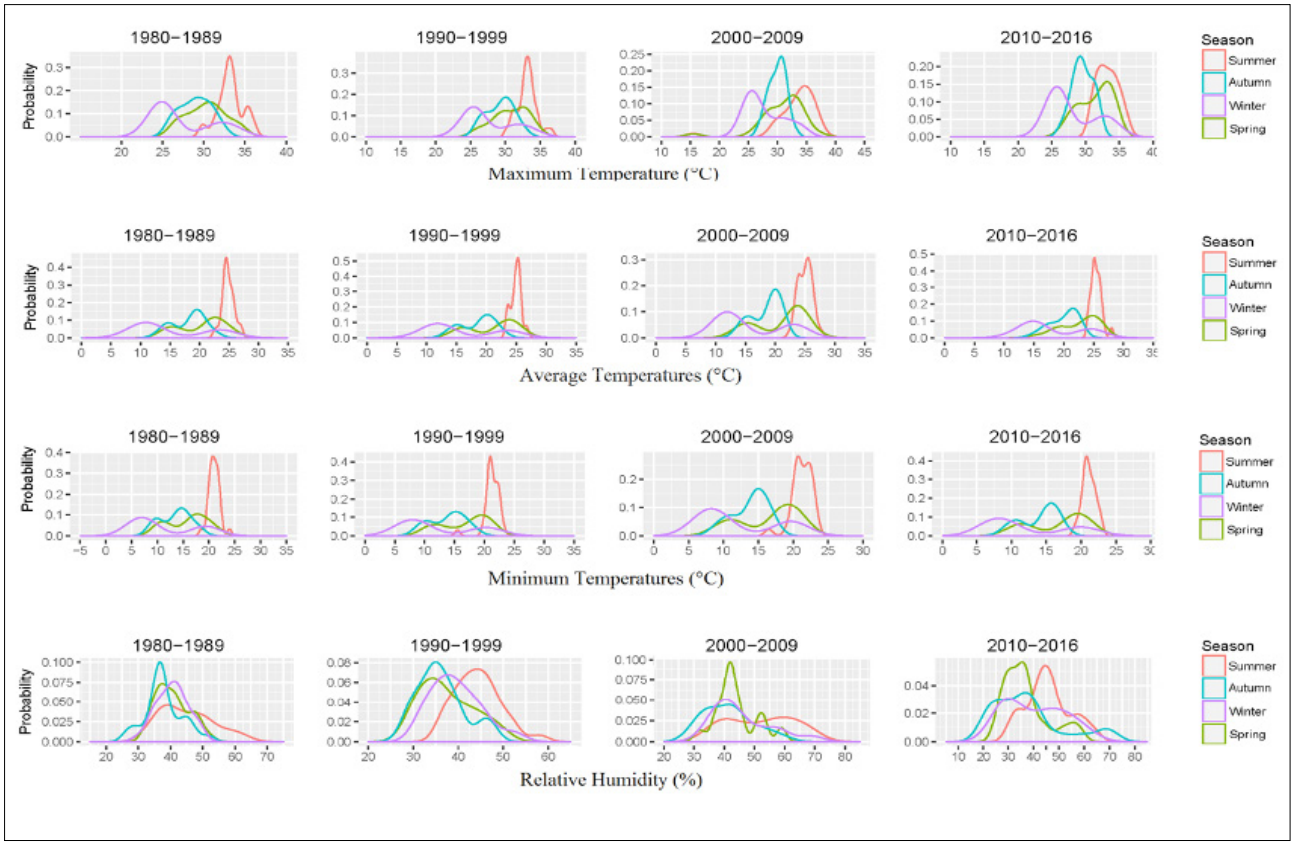



**Figure 9:** Decadal probability densities of seasonal relative humidity and air temperature (maximum, average and minimum) trends for northernmost Limpopo Province, 1980–2016. Probability density plots, which establish variability of relative humidity and air temperature, were calculated using R software package ggplot2.

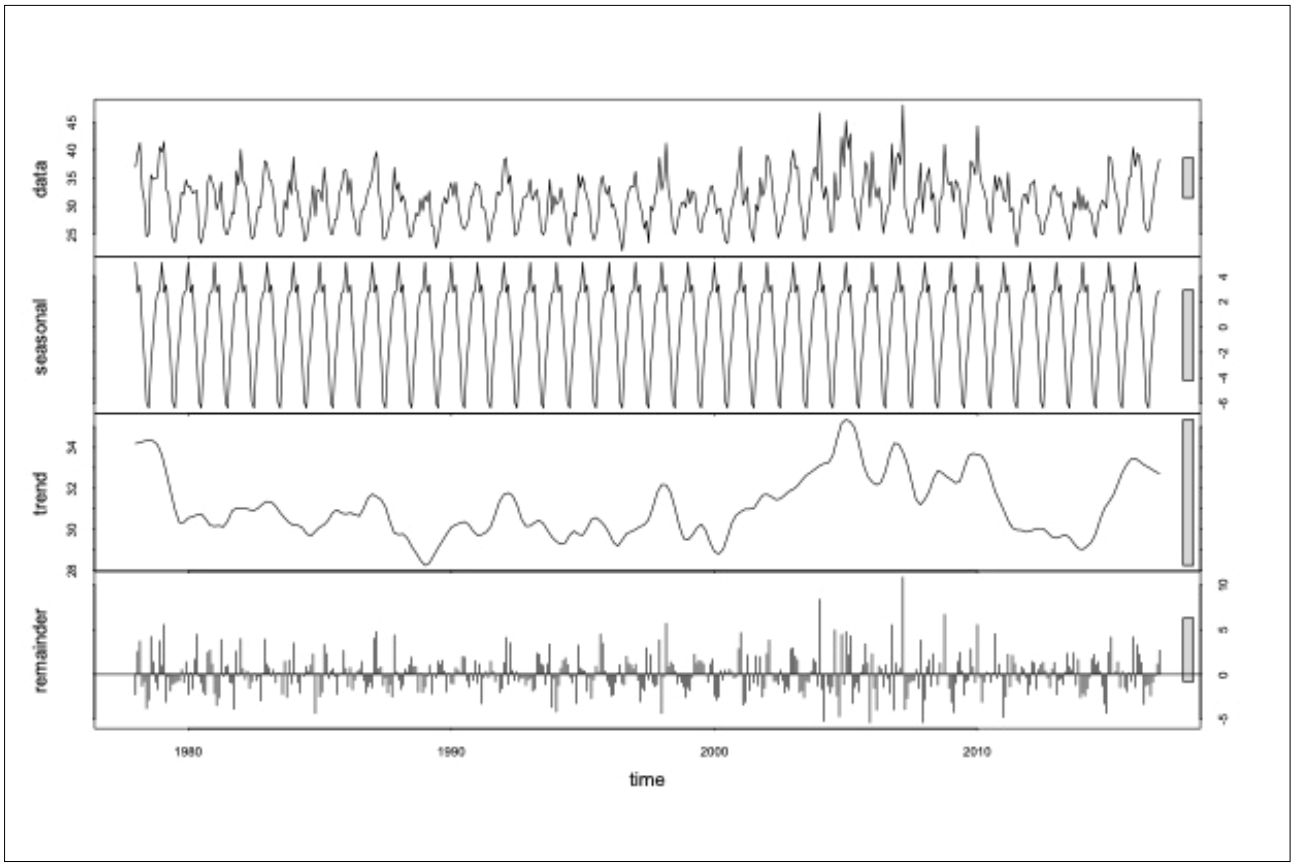



**Figure 10:** Seasonal and trend decomposition of the heat index (°C) for northernmost Limpopo Province, 1980–2016. 'Data' = set of actual values; 'seasonal' component = period; 'trend' = increasing or decreasing direction in the data; and 'remainder' = residual. The grey bar at the right-hand side of each graph allows for a relative comparison of the magnitudes of each component.

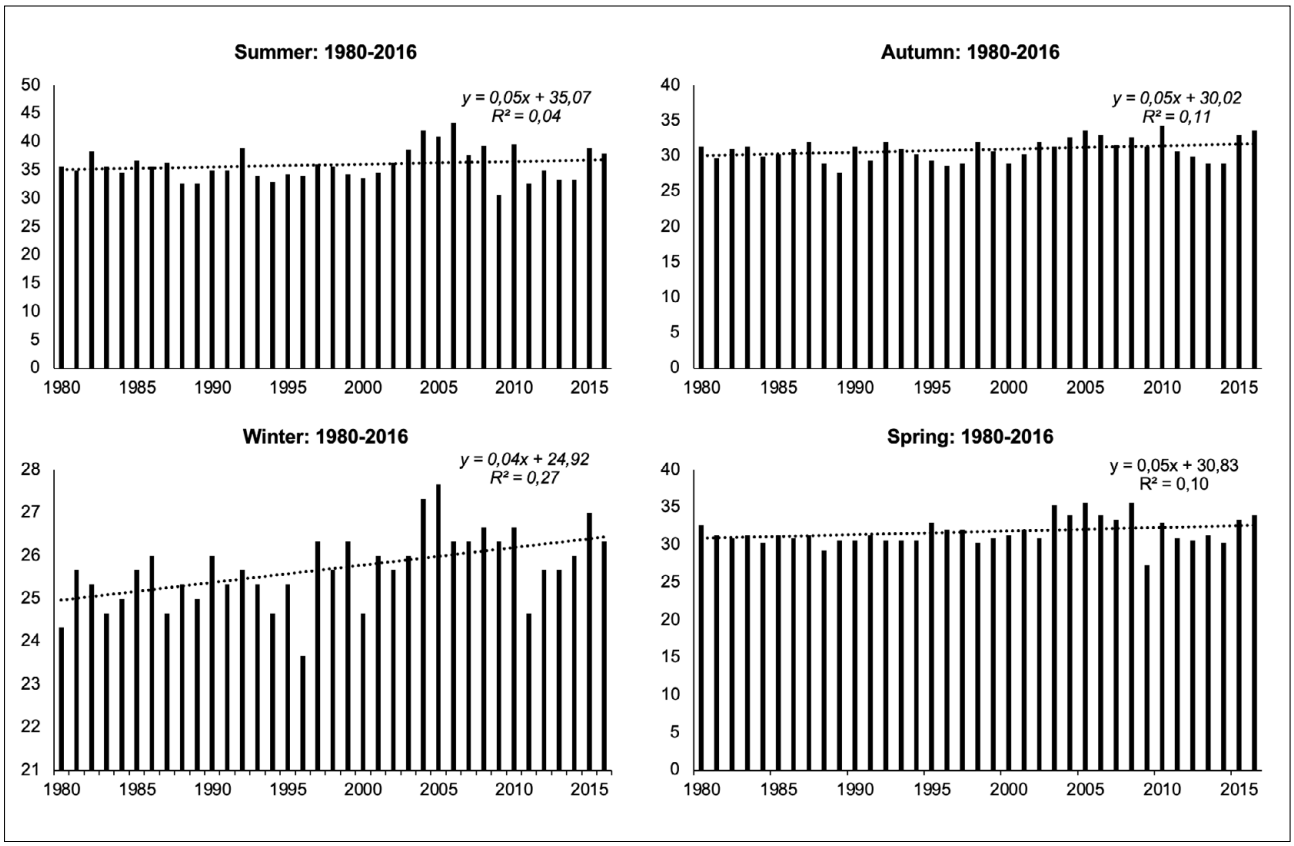



**Figure 11:** Seasonal heat index values based on maximum temperatures and relative humidity for the period 1980–2016 in northernmost Limpopo Province.

Increases in the frequency, occurrence, intensity and amplitude of heatwaves, as well as positive trends in the heat index (i.e. how hot it feels) have direct adverse effects on agriculture, especially livestock production. High temperatures cause heat stress in livestock.[40] For instance, in poultry production, ambient temperatures above 30 °C are known to cause heat stress in chickens, resulting in reduced feed intake and body weight, and, in some cases, high mortality.[4,41] This is a cause for concern in rural communities of northernmost Limpopo Province, which are resource limited, with low adaptive capacities. A recent study by Mbokodo et al.[28] focusing on expected heatwaves in a future warmer South Africa, projects an increased frequency and duration of heatwaves. Such a scenario carries the likely consequence of significant adverse impacts on human health, economic activities (especially climate-sensitive farming) and livelihoods in vulnerable communities such as the northern border region of Limpopo Province.

Both minimum and maximum temperatures increased at a similar rate (+0.02 °C/year) over the study period – a finding inconsistent with previous studies in the region and elsewhere, which reported more rapid increases in maximum temperature compared to minimum temperatures.[14,33,42,43] The observed rates of increase in minimum and maximum temperatures in northernmost Limpopo Province are inconsistent with several global warming reports, which have shown a decreasing diurnal temperature range due to minimum temperatures increasing at a faster rate than maximum temperatures.[42-46] Diurnal temperature range is an important indicator of climate change, and is influenced by various spatial and temporal factors, including land-use / land-cover changes, irrigation, station moves, desertification, and a host of other indirect climatic effects.[47] Thus, although global and regional perspectives on changes in the climate system are important, there are complex spatial variations, which necessitate local analysis of trends.

Simulated climate models (e.g. regional climate models and general circulation models) are often used for climate change projections. However, models may have limitations in determining local weather conditions.[13,33] It is thus important to test and refine model reports using observed variations to ensure the accuracy of climate change projections. For instance, Jury[13] analysed spatial and temporal historical climate observations in southern Africa to validate simulation models for projecting climate trends. Kruger et al.[48] investigated historical trends in near-surface minimum and maximum temperatures, as well as extreme temperature indices in South Africa. This was achieved through critically comparing quality-controlled station observations with downscaled model projections, so as to provide valuable information concerning the interpretation of model-generated projections. Kruger et al.'s[48] results demonstrate that model outputs tend to simulate the historical trends accurately for annual means of daily maximum and minimum temperatures, but have limitations when assessing temperature extremes. For correctly estimating the potential impact of climate change in a given region, particularly for sectors such as agriculture, it is necessary to reconcile climate observations and model projections.

Climate change research in remote rural areas such as the northern region of Limpopo Province should be ongoing, especially given that trends of concern seem to be increasing. Thus, climate change has an ever-changing impact on farming, especially on small-scale farming systems and enterprises in the region. Information on climatic trends, especially at local level, is important to help understand the likely impacts on farming productivity, and to inform appropriate strategic interventions to improve adaptation mechanisms.[47] It is likely that temperatures in northernmost Limpopo Province are frequently surpassing livestock heat stress thresholds each year, ultimately compromising productivity, yet farmers continue with agricultural production without such insights. It is possible that the reported low productivity of rural poultry in the northern region of Limpopo Province may, in part, be a function of heat stress.[41] Thus, access to climate information is essential to provide understanding of likely implications of climate change on agricultural production – something that needs to be effectively communicated to small-scale subsistence farmers for appropriate adaption.

## Conclusion

Temperatures in northernmost Limpopo Province have steadily increased over the period 1950–2016. Consequently, heatwave frequency, intensity and amplitude have also increased. Further, temperatures and relative humidity have become more variable in recent decades. Of particular concern is the fact that, according to heat index trends, heat indices have increased. Such increases are likely to have adverse effects on sectors such as agriculture, especially in small-scale farming systems where farmers have limited adaptive capacities. To establish appropriate preparedness, an important starting point would be for future research to focus on establishing heat-stress tolerance thresholds for specific agricultural sectors in the northern border region of Limpopo Province.

## Acknowledgements

We acknowledge the financial assistance from Open Society Foundation through the University of the Witwatersrand Global Change Institute as well as the weather data provided by the South African Weather Service.

## Competing interests

We have no competing interests to declare.

## Authors' contributions

N.M.B.N. was the principal researcher and main author of the manuscript. S.G. was the main supervisor of the research work and reviewed the manuscript. J.M. and E.A. co-supervised the research, analysed the data and reviewed the manuscript.

## References

1. Stocker TF, Qin D, Plattner GK, Tignor M, Allen SK, Boschung J, et al. Summary for policymakers. In: Climate change 2013: The physical science basis: Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge: Cambridge University Press; 2014. https://doi.org/10.1017/cbo9781107415324.004
2. Ziervogel G, New M, Archer van Garderen E, Midgley G, Taylor A, Hamann R, et al. Climate change impacts and adaptation in South Africa. WIREs Clim Change. 2014;5(5):605–620. https://doi.org/10.1002/wcc.295
3. Van Wilgen NJ, Goodall V, Holness S, Chown SL, McGeoch MA. Rising temperatures and changing rainfall patterns in South Africa's national parks. Int J Climatol. 2016;36(2):706–721. https://doi.org/10.1002/joc.4377
4. South African Department of Environmental Affairs. Long-Term Adaptation Scenarios Flagship Research Programme (LTAS) for South Africa. Climate trends and scenarios for South Africa [document on the Internet]. c2013 [cited 2017 Mar 04]. Available from: https://www.sanbi.org/sites/default/files/documents/documents/ltasclimate-trends-and-scenarios-tech-report2013low-res.pdf
5. Thornton PK, Ericksen PJ, Herrero M, Challinor AJ. Climate variability and vulnerability to climate change: A review. Glob Change Biol. 2014;20(11):3313–3328. https://doi.org/10.1111/gcb.12581
6. Diallo I, Giorgi F, Sukumaran S, Stordal F, Giuliani G. Evaluation of RegCM4 driven by CAM4 over Southern Africa: Mean climatology, interannual variability and daily extremes of wet season temperature and precipitation. Theor Appl Climatol. 2015;121(3–4):749–766. https://doi.org/10.1007/s00704-014-1260-6
7. Engelbrecht F, Adegoke J, Bopape MJ, Naidoo M, Garland R, Thatcher M, et al. Projections of rapidly rising surface temperatures over Africa under low mitigation. Environ Res Lett. 2015;10(8):1–16. https://doi.org/10.1088/1748-9326/10/8/085004
8. Kusangaya S, Warburton ML, Archer van Garderen ERM, Jewitt GP. Impacts of climate change on water resources in southern Africa: A review. Phys Chem Earth. 2014;67–69:47–54. http://dx.doi.org/10.1016/j.pce.2013.09.014
9. Keggenhoff I, Elizbarashvili M, King L. Recent changes in Georgia's temperature means and extremes: Annual and seasonal trends between 1961 and 2010. Weather Clim Extrem. 2015;8:34–45. https://doi.org/10.1016/j.wace.2014.11.002

10. Hachigonta S, Nelson GC, Thomas TS, Sibanda LM, editors. Southern African agriculture and climate change: A comprehensive analysis. Washington DC: International Food Policy Research Institute; 2013. http://dx.doi.org/10.2499/9780896292086

11. Tadross M, Jack C, Hewitson B. On RCM-based projections of change in southern African summer climate. Geophys Res Lett. 2005;32(23):1–4. https://doi.org/10.1029/2005gl024460

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

13. Jury MR. Climate trends in southern Africa. S Afr J Sci. 2013;109(1–2), Art. #980. http://dx.doi.org/10.1590/sajs.2013/980

14. Kruger AC, Sekele SS. Trends in extreme temperature indices in South Africa: 1962–2009. Int J Climatol. 2013;33(3):661–676. https://doi.org/10.1002/joc.3455

15. Gbetibouo GA. Understanding farmers' perceptions and adaptations to climate change and variability: The case of the Limpopo Basin, South Africa. Washington DC: International Food Policy Research Institute; 2009. https://www.ifpri.org/publication/understanding-farmers-perceptions-and-adaptations-climate-change-and-variability

16. Tshiala MF, Olwoch JM, Engelbrecht FA. Analysis of temperature trends over Limpopo province, South Africa. J Geog Geol. 2011;3(1):13. https://doi.org/10.5539/jgg.v3n1p13

17. Kalumba AM, Olwoch JM, Van Aardt I, Botai OJ, Tsela P, Nsubuga FW, et al. Trend analysis of climate variability over the West Bank-East London Area, South Africa (1975-2011). J Geog Geol. 2013;5(4):131. https://doi.org/10.5539/jgg.v5n4p131

18. Lakhraj-Govender R, Grab S, Ndebele NE. A homogenized long-term temperature record for the Western Cape Province in South Africa: 1916-2013. Int J Climatol. 2017;37(5):2337–2353. https://doi.org/10.1002/joc.4849

19. Archer van Garderen ER. (Re) considering cattle farming in southern Africa under a changing climate. Weather Clim Soc. 2011;3(4):249–253. https://doi.org/10.1175/wcas-d-11-00026.1

20. Shisanya S, Mafongoya P. Adaptation to climate change and the impacts on household food security among rural farmers in uMzinyathi District of Kwazulu-Natal, South Africa. Food Secur. 2016;8(3):597–608. https://doi.org/10.1007/s12571-016-0569-7

21. Makhado R, Potgieter M, Luus-Powell W, Cooper S, Oppong C, Kopij G, et al. *Tragelaphus strepsiceros* browse during the dry season in the mopani veld of Limpopo Province, South Africa. Trans R Soc S Afr. 2016;71(1):17–21. https://doi.org/10.1080/0035919x.2015.1102174

22. Tankson JD, Vizzier-Thaxton Y, Thaxton JP, May JD, Cameron JA. Stress and nutritional quality of broilers. Poult Sci J. 2001;80:1384–1389. https://doi.org/10.1093/ps/80.9.1384

23. Boissonnade AC, Heitkemper LJ, Whitehead D. Weather data: Cleaning and enhancement. In: Dischel RS. Climate risk and the weather market: Financial risk management with weather hedges. London: Risk Books; 2002. p. 73–93. Available from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.527.2282&rep=rep1&type=pdf

24. Vincent LA, Wang XL, Milewska EJ, Wan H, Yang F, Swail V. A second generation of homogenized Canadian monthly surface air temperature for climate trend analysis. J Geophys Res Atmos. 2012;117(D18):1–13. https://doi.org/10.1029/2012jd017859

25. Wang XL, Feng Y, Vincent LA. Observed changes in one-in-20-year extremes of Canadian surface air temperatures. Atmos Ocean. 2014;52(3):222–231. https://doi.org/10.1080/07055900.2013.818526

26. Spinu V, Grolemund G, Wickham H. lubridate: Make dealing with dates a little easier. R package version 1.7.4; 2018.

27. Hyndman RJ, Athanasopoulos G, Bergmeir C, Caceres G, Chhay L, O'Hara-Wild M, et al. Package 'forecast'; 2020. Available from: https://cran.r-project.org/web/packages/forecast/forecast.pdf

28. Mbokodo I, Bopape MJ, Chikoore H, Engelbrecht F, Nethengwe N. Heatwaves in the future warmer climate of South Africa. Atmosphere. 2020;11(7):712. https://doi.org/10.3390/atmos11070712

29. Wickham H. ggplot2. WIREs Comp Stat. 2011;3(2):180–185. https://doi.org/10.1002/wics.147

30. Anderson GB, Peng RD. weathermetrics: Functions to convert between weather metrics (R package). Version 1.2.2. 2016.

31. National Weather Service Hydrometeorological Prediction Center Web Team. Heat Index Calculator; 2015. Available from: https://www.wpc.ncep.noaa.gov/html/heatindex.shtml

32. Zhang X, Feng Y, Chan R. Introduction to RClimDex v1.9. Environment Canada; 2018. Available from: https://usermanual.wiki/Document/manual.2056401896/view

33. Davis CL, Hoffman MT, Roberts W. Recent trends in the climate of Namaqualand, a megadiverse arid region of South Africa. S Afr J Sci. 2016;112(3–4), Art. #2015-0217. https://doi.org/10.17159/sajs.2016/20150217

34. Kruger AC, Nxumalo M. Surface temperature trends from homogenized time series in South Africa: 1931-2015. Int J Climatol. 2017;37(5):2364–2377. https://doi.org/10.1002/joc.4851

35. Huth R, Kyselý J, Pokorná L. A GCM simulation of heat waves, dry spells, and their relationships to circulation. Clim Change. 2000;46(1–2):29–60. https://doi.org/10.1023/a:1005633925903

36. Meehl GA, Tebaldi C. More intense, more frequent, and longer lasting heat waves in the 21st century. Science. 2004;305(5686):994–997. https://doi.org/10.1126/science.1098704

37. Schlenker W, Roberts MJ. Nonlinear temperature effects indicate severe damages to US crop yields under climate change. Proc Natl Acad Sci USA. 2009;106(37):1594–1598. https://doi.org/10.1073/pnas.0906865106

38. Van der Velde M, Tubiello FN, Vrieling A, Bouraoui F. Impacts of extreme weather on wheat and maize in France: Evaluating regional crop simulations against observed data. Clim Change. 2012;113(3–4):751–765. https://doi.org/10.1007/s10584-011-0368-2

39. Lefcourt AM, Adams WR. Radiotelemetry measurement of body temperatures of feedlot steers during summer. J Anim Sci. 1996;74(11):2633–2640. https://doi.org/10.2527/1996.74112633x

40. Thornton PK, Van de Steeg J, Notenbaert A, Herrero M. The impacts of climate change on livestock and livestock systems in developing countries: A review of what we know and what we need to know. Agric Syst. 2009;101(3):113–127. https://doi.org/10.1016/j.agsy.2009.05.002

41. Nyoni NMB, Grab S, Archer ER. Heat stress and chickens: climate risk effects on rural poultry farming in low-income countries. Clim Dev. 2019;11(1):83–90. https://doi.org/10.1080/17565529.2018.1442792

42. Easterling DR, Horton B, Jones PD, Peterson TC, Karl TR, Parker DE, et al. Maximum and minimum temperature trends for the globe. Science. 1997;277(5324):364–367. https://doi.org/10.1126/science.277.5324.364

43. New M, Hewitson B, Stephenson DB, Tsiga A, Kruger A, Manhique A, et al. Evidence of trends in daily climate extremes over southern and west Africa. J Geophys Res Atmos. 2006;111(D14):1–11. https://doi.org/10.1029/2005jd006289

44. Karl TR, Jones PD, Knight RW, Kukla G, Plummer N, Razuvayev V, et al. A new perspective on recent global warming: Asymmetric trends of daily maximum and minimum temperature. Bull Am Meteor Soc. 1993;74(6):1007–1024. https://doi.org/10.1175/1520-0477(1993)074<1007:anporg>2.0.co;2

45. Liu B, Xu M, Henderson M, Qi Y, Li Y. Taking China's temperature: Daily range, warming trends, and regional variations, 1955-2000. J Clim. 2004;17(22):4453–4462. https://doi.org/10.1175/3230.1

46. Blunden J, Arndt DS. State of the climate in 2011. Warming: Asymmetric trends of daily maximum and minimum temperature. Bull Am Meteor Soc. 2012;93(7):S1–282. https://doi.org/10.1175/3230.1

47. Qu M, Wan J, Hao X. Analysis of diurnal air temperature range change in the continental United States. Weather Clim Extrem. 2014;4:86–95. https://doi.org/10.1016/j.wace.2014.05.002

48. Kruger AC, Rautenbach H, Mbatha S, Ngwenya S, Makgoale TE. Historical and projected trends in near-surface temperature indices for 22 locations in South Africa. S Afr J Sci. 2019;115(5–6), Art. #4846. https://doi.org/10.17159/sajs.2019/4846

**AUTHORS:**
Patrick Nyamaruze[1]
Kaymarlin Govender[2]
Richard G. Cowden[3]

**AFFILIATIONS:**
[1]Discipline of Psychology, School of Applied Human Sciences, University of KwaZulu-Natal, Durban, South Africa
[2]Health Economics and HIV and AIDS Research Division, University of KwaZulu-Natal, Durban, South Africa
[3]Department of Psychology, University of the Free State, Bloemfontein, South Africa

**CORRESPONDENCE TO:**
Patrick Nyamaruze

**EMAIL:**
nyamruzepatrick@yahoo.com

**DATES:**
**Received:** 18 May 2020
**Revised:** 09 July 2021
**Accepted:** 22 July 2021
**Published:** 29 Nov. 2021

**HOW TO CITE:**
Nyamaruze P, Govender K, Cowden RG. Self-esteem and antiretroviral therapy adherence among young people living with HIV: An exploratory serial mediation analysis. S Afr J Sci. 2021;117(11/12), Art. #8354. https://doi.org/10.17159/sajs.2021/8354

**ARTICLE INCLUDES:**
☒ Peer review
☐ Supplementary material

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

**EDITORS:**
Pascal Bessong
Sandiswa Mbewana

**KEYWORDS:**
adolescents, young adults, HIV, HIV/AIDS stigma, body perceptions, medication adherence

**FUNDING:**
None

# Self-esteem and antiretroviral therapy adherence among young people living with HIV: An exploratory serial mediation analysis

Capitalising further on the benefits of antiretroviral therapy (ART) for individual treatment requires an improved understanding of the psychological processes that may affect optimal ART adherence among people living with HIV. We examined internalised HIV/AIDS-related stigma and body appreciation as mediators of the association between self-esteem and ART adherence among young people living with HIV (YPLHIV). A sample of 76 YPLHIV ($M_{age}$ = 19.36, $s.d._{age}$ = 2.56; male 56.58%) residing in an HIV hyperendemic region of South Africa completed self-report measures of self-esteem, internalised HIV/AIDS-related stigma, body appreciation, and ART adherence. Path-analytic mediation modelling was performed to test for direct and indirect effects linking self-esteem with ART adherence. Results of serial mediation analyses indicated that self-esteem and ART adherence were indirectly associated through a two-step path of internalised HIV/AIDS-related stigma and then body appreciation, as well as a one-step path through internalised HIV/AIDS-related stigma. The results provide preliminary support for internalised HIV/AIDS-related stigma and body appreciation as mechanisms underlying the association between self-esteem and ART adherence. Implications of the findings for promoting ART adherence among YPLHIV are discussed.

**Significance:**

- Self-esteem and ART adherence were indirectly related through internalised HIV/AIDS-related stigma followed by body appreciation.
- Outcomes of intervention initiatives designed to promote ART adherence among young people living with HIV may be further improved by integrating components that target internalised HIV/AIDS-related stigma and body appreciation.

## Introduction

Increased availability and uptake of antiretroviral therapy (ART) among people living with HIV (PLHIV) have played an important role in the decline of new HIV infections.[1] ART promotes viral suppression, improves physical and emotional well-being, and reduces mortality.[2] In regions of HIV hyperendemicity where it has been particularly challenging to stay on track with global treatment and prevention targets, ART uptake represents a key avenue for reducing the burden of HIV. However, the benefits of ART for preventing and treating HIV may be undermined by treatment non-adherence, which can lead to impairments in psychosocial functioning (e.g. weakening of social relationships and loss of self-esteem resulting from periods of illness), increase risk of HIV transmission to non-infected sexual partners, and is a common cause of death among PLHIV.[3,4] To advance health promotion initiatives designed to improve ART adherence, further research is needed to identify and better understand salient mechanisms underlying ART non-adherence in populations residing in HIV hyperendemic contexts and for which unique social-structural vulnerabilities exist.

The burden of the global HIV epidemic is highest among young people living in the countries within eastern and southern Africa.[5] Alongside the social-structural issues (e.g. inadequate health infrastructure) that pose challenges to HIV testing and treatment-related behaviour of young people living in the eastern and southern African region[6], the effectiveness of HIV treatment in promoting the health and longevity of this population is highly contingent on the individual choices young people make about treatment-related behaviour. Wide-ranging reasons for non-adherence to ART have been reported among young people. The more frequently identified patient-related factors include perceived or internalised HIV/AIDS-related stigma[7], body image concerns[8], forgetfulness, alcohol use, and depression[9]. Patient-related factors are closest in proximity to the individual, many of which appear amenable to change through targeted efforts that are directed towards harnessing and developing personal resources of young people living with HIV (YPLHIV). Although various psychological resources (e.g. self-efficacy, resilience) have been linked to better ART adherence[10,11], one underexplored avenue that also has the potential to promote ART adherence among YPLHIV is self-esteem.

Self-esteem is an important inner resource encompassing an individual's overall positive evaluation of the self, including perceived competence and sense of worth.[12] Self-esteem has been conceptualised as a protective factor that contributes to better health and well-being by buffering the impact of negative and stressful life experiences.[13,14] Research involving young people has revealed that self-esteem is associated with lower health risk behaviour (e.g. fewer sexual partners[15], lower alcohol use[16]), which suggests that those with higher self-esteem might be more inclined to make choices that support rather than degrade health and well-being (e.g. adhere to medication regimens).

Few studies have investigated the association between self-esteem and ART adherence, which is important for several reasons. First, negative self-evaluation and a low sense of self-worth may affect concentration and memory, both of which are associated with ART non-adherence.[17] For example, impairment in prospective memory (i.e. remembering to perform a specific task in the future) has been associated with ART non-adherence.[18] Second, lower self-esteem is linked to increased alcohol misuse and recreational substance use, which can impair judgement and may interfere with a person's ability to adhere to treatment.[16] Third, high self-esteem can precipitate resilience-promoting psychological qualities (e.g. self-confidence) that may enhance medication adherence.[13] To further examine the association between self-esteem and ART adherence, in the present study we tested for evidence of internalised HIV/AIDS-related stigma and body appreciation as serial mechanisms linking self-esteem and ART adherence in a sample of South African YPLHIV.

### *Self-esteem and internalised HIV/AIDS-related stigma*

Internalised HIV/AIDS-related stigma refers to the negative perceptions and abasement of the self that may be endorsed by those diagnosed with HIV.[19] It is characterised by self-denigrating thoughts (e.g. self-blame) and feelings (e.g. shame, guilt), as well as concealment of HIV status.[20] Internalisation of stigma is not an inevitable response to possession of a stigmatised attribute[21], but the risk of such increases when existing beliefs about one's self-worth and self-regard are low. In particular, low self-esteem is associated with a higher likelihood of acquiring identity standards from society, appropriating perspectives about the self from others, and evaluating the self negatively.[22]

Although few studies have directly investigated the impact of self-esteem on internalised HIV/AIDS-related stigma, existing research indicates that self-esteem may protect against the negative effects that perceived or experienced stigma can have on internalisation of HIV/AIDS-related stigma.[23] Some findings suggest that positive changes in self-esteem (e.g. increased self-worth) can play a role in adaptive resistance to HIV/AIDS-related stigma (e.g. open disclosure of HIV seropositive status to confront negative labelling).[24] Other evidence points to the risk of internalising HIV/AIDS-related stigma when self-esteem is low, which could have negative downstream implications for health and well-being.[25,26]

### *Internalised HIV/AIDS-related stigma and body appreciation*

Internalised HIV/AIDS-related stigma may be pronounced when PLHIV have physical attributes that are common markers of the disease.[27] Previous studies have established a link between internalised HIV/AIDS-related stigma and body image[28], with one study reporting that high stigma significantly lowered the probability of having a positive body image[29]. Even in the absence of visible body changes, self-deprecating psychological processes that are attributable to internalised HIV/AIDS-related stigma (e.g. feelings of shame and worthlessness) can impact negatively on the body image perceptions of PLHIV.[30]

Internalised HIV/AIDS-related stigma may affect different dimensions of body image, including body appreciation. Body appreciation is a component of positive body image that is conceptualised as being respectful, approving, and appreciative of the characteristics and health of one's body, regardless of shape, weight, and imperfections.[31] Internalising social norms and discriminatory characterisations towards PLHIV may have a negative effect on how PLHIV view and appreciate their bodies.[27] The negative implications of internalised HIV/AIDS-related stigma may be especially pronounced among YPLHIV, given the complex interplay of biopsychosocial processes that occur during the developmental transition from childhood to adulthood.[6]

### *Body appreciation and antiretroviral therapy adherence*

A variety of physical and psychological complications commonly accompany HIV infection, including weight loss, skin lesions, depression, suicidal ideation, and body dissatisfaction.[32,33] Research suggests that fear of bodily changes (e.g. weight loss) resulting from HIV and ART may be implicated in medication non-adherence.[8] Although there are few direct links between body appreciation and adherence, there is some evidence that improving the perspectives of PLHIV about their body can lead to better health behaviour, including adherence to ART.[34] Extensive evidence indicates that body appreciation is associated with an appreciation of diverse appearances and shapes[31], better physical health, and higher levels of psychological well-being[35]. Efforts to promote body appreciation may have implications for ART adherence among PLHIV, especially among young people who tend to be particularly susceptible to the influence of bodily perceptions on positive health behaviour.[36]

## The present study

In this study, we examine associations between self-esteem, internalised HIV/AIDS-related stigma, body appreciation, and ART adherence in YPLHIV who reside in an HIV hyperendemic country within the eastern and southern African region. Exploring these interrelationships could offer an improved understanding of mechanisms that are implicated in ART adherence, particularly those that could be targeted to promote adherence in seropositive young people. Thus, we used an integrative modelling approach to explore internalised HIV/AIDS-related stigma and body appreciation as serial mechanisms linking self-esteem and ART adherence among YPLHIV in the Durban area of KwaZulu-Natal, South Africa. To our knowledge, this is the first study examining body appreciation as a mediator of this association, particularly as a precursor to ART adherence. We hypothesised that self-esteem and ART adherence would be linked indirectly via a sequential path of internalised HIV/AIDS-related stigma followed by body appreciation.

## Method

### *Participants*

The sample comprised 76 (56.58% male) seropositive young people residing in the city of Durban, South Africa. Durban is located within the province of KwaZulu-Natal, which has the highest HIV prevalence rate in the country.[37] Participants ranged from 15 to 24 years of age ($M_{age}$=19.36, s.d.=2.56) and identified racially as African (82.89%) or coloured (17.11%). A majority of the sample had fulfilled high school equivalency requirements (67.11%) or completed post-secondary education (28.95%). A small proportion of the participants had not completed any formal education (3.95%). Almost half of the sample (52.63%) indicated that they had acquired HIV after birth from a source other than their mother (i.e. horizontal infection), and the remainder (47.37%) reported that they had been infected through mother-to-child transmission (i.e. vertical infection).

### *Measures*

#### Rosenberg Self-Esteem Scale

The Rosenberg Self-Esteem Scale (RSES)[12] is a 10-item measure of self-respect and self-acceptance. Items (e.g. 'I feel that I am a person of worth, at least on an equal plane with others') are rated using a four-point response scale (1 = Strongly disagree; 4 = Strongly agree), half of which are reverse scored. In this study, responses to all items were aggregated for a total raw score (range: 10 to 40). Findings of various studies support the construct validity of the RSES.[38,39] Estimated internal consistency reported for the RSES in prior research has been ≥0.80.[40]

#### Internalised AIDS-Related Stigma Scale

The Internalised AIDS-Related Stigma Scale (IA-RSS)[41] consists of six items that were adapted from the AIDS-Related Stigma Scale[42] to measure negative self-perceptions and self-abasement about being a person living with HIV. Items (e.g. 'I hide my HIV status from others') are rated on a dichotomous response scale (0 = Disagree; 1 = Agree) and are summed for a total score ranging from 0 to 6. Higher scores indicate greater internalised HIV/AIDS-related stigma. Support for the construct validity of the IA-RSS has been evidenced through associations with indicators of mental health and well-being (e.g. depression, quality of life).[19] Prior research involving samples of South Africans have revealed internal consistency values of ≥0.73 for the IA-RSS.[41]

### Body Appreciation Scale-2

The Body Appreciation Scale-2 (BAS-2)[43] contains 10 items that measure a person's perceptions of their body (e.g. acceptance, appreciation, inner positivity) and attention towards their body's needs through the adoption of healthy behaviours.[43] Participants rate the items (e.g. 'I feel good about my body') using a five-point response scale (1 = Never; 5 = Always). Item responses are summed for a total score ranging from 10 to 50. Higher scores indicate greater levels of body appreciation. The BAS-2 has been cross-culturally validated in samples from diverse countries (e.g. China, England).[44,45] Previous studies have reported internal consistency values of ≥0.80 for women and men[44], and evidence supports the convergent, incremental, and discriminant validity of the BAS-2[43].

### Morisky Medication Adherence Questionnaire

The Morisky Medication Adherence Questionnaire (MMAS-8)[46] comprises eight items that assess patient adherence to medication for chronic conditions. The first seven items (e.g. 'When you feel like your symptoms are under control, do you sometimes stop taking your medicine') are rated using a dichotomous response format (0 = Yes; 1 = No). A five-point response format is used to rate the final item (e.g. 'How often do you have difficulty remembering to take all your medicine'), which is transformed to a dichotomous score prior to aggregation of item responses (0 = A; 1 = B–E). In this study, participants completed the MMAS-8 by referencing their adherence to ART medication. We aggregated responses to each of the MMAS-8 items for a total score ranging from 0 to 8, with higher scores reflecting greater adherence to ART. Evidence supports the psychometric utility of the MMAS-8 as a valid and reliable measure of medication adherence[46], including ART adherence[47]. Previous research has reported internal consistency values of ≥0.75 for the MMAS-8, and the findings of several studies support the construct validity of the measure.[46,48]

## *Procedure*

Ethical approval for this study was granted by the Humanities and Social Sciences Research Ethics Committee of the University of KwaZulu-Natal (HSS/0522/018D). Written permission was acquired from a local HIV youth centre to access YPLHIV who received ART from the organisation. A purposive, convenience sampling approach was used to recruit participants. Specifically, YPLHIV who presented at the youth centre to collect their monthly medication were invited by a team of researchers to participate in this study. Interested individuals were directed to a private administrative office where they were initially given details about the study purpose, the participation procedures, and ethical considerations (e.g. anonymity, confidentiality). Those who agreed to participate were invited to provide their written informed consent. With the assistance of the nurses at the youth centre, parents and legal guardians of potential participants below 18 years were informed of the research. Written parental consent was obtained on behalf of all legal minors who indicated their interest in participating. Written assent was also obtained from legal minors. The measures were administered in English, which participants completed in an office provided by the youth centre. The research team was present and available to address any questions or issues that arose during the course of the subjects' participation.

## Results

Statistical processing was performed using R.[49] Study variables were initially screened for gross univariate and multivariate outliers. Standardised values for all variables were within acceptable limits (i.e. $z \leq |3.29|$)[50], indicating there were no univariate outlier concerns. Mahalanobis distance ($\chi^2$ (4) = 18.47, $p<0.001$) did not reveal any multivariate outliers (all $D^2$ values ≤11.92). Univariate skewness (max. = |0.86|) and kurtosis (max. = |0.76|) values indicated that all variables were approximately normal in distribution. Descriptive statistics, internal consistency estimates, and zero-order correlations among the study variables are reported in Table 1. Omega total ($\omega_t$) estimates of internal consistency for all measures were ≥0.74. Self-esteem evidenced a small negative association with internalised HIV/AIDS-related stigma and a small positive association with body appreciation, but its association with ART adherence was negligible. Moderate negative associations were found linking internalised HIV/AIDS-related stigma with both body appreciation and ART adherence. There was a moderate positive association between body appreciation and ART adherence.

Path modelling procedures were performed using maximum likelihood estimation with robust standard errors. Evidence of mediation was tested using a global approach, which involves determining mediation effects after first establishing adequacy of model fit. Following existing recommendations[51,52], model fit was evaluated using a combination of absolute, incremental, and residual-based fit indices. Specifically, we report the chi-square goodness-of-fit statistic (and its associated *p*-value), the comparative fit index, the Tucker Lewis index, the standardised root mean square residual, and the root mean square error of approximation. Along with the statistical significance of the chi-square test statistic ($p>0.05$), values of ≥0.90 for comparative fit index and Tucker Lewis index and values ≤0.10 for root mean square error of approximation and standardised root mean square residual were used as benchmarks to guide our evaluation of model fit.[53,54] We also report the Akaike information criterion to allow comparisons between estimated models, with lower Akaike information criterion values indicative of a more favourable level of fit.[55] Based on our theorising that internalised HIV/AIDS-related stigma and body appreciation would be serial mediators linking self-esteem with medication adherence, we began with the most parsimonious baseline full mediation model in which the association between self-esteem (X) and ART adherence (Y) was specified to occur via internalised HIV/AIDS-related stigma ($M_1$) and then body appreciation ($M_2$). Additional paths were sequentially integrated into subsequent models to determine whether less restrictive models were favoured over those that were more parsimonious. Because model estimation was based on cross-sectional data, a series of conceptually viable alternative models was also estimated.

**Table 1:** Descriptive statistics, internal consistency estimates, and zero-order correlations among study variables (*N*=76)

| Variable | Mean ± s.d. (observed range) | (1) | (2) | (3) | (4) |
|---|---|---|---|---|---|
| (1) Self-esteem | 15.61 ± 1.63 (11, 19) | (0.82) | | | |
| (2) Internalised HIV/AIDS-related stigma | 2.93 ± 1.37 (1, 6) | -0.23** [-0.43, -0.01] | (0.74) | | |
| (3) Body appreciation | 40.49 ± 5.20 (29, 50) | 0.20* [-0.02, 0.41] | -0.40*** [-0.57, -0.19] | (0.84) | |
| (4) Antiretroviral therapy adherence | 5.17 ± 1.59 (1, 8) | 0.06 [-0.17, 0.28] | -0.36** [-0.54, -0.15] | 0.44*** [0.24, 0.61] | (0.77) |

*p<*0.10;* **p<*0.05;* ***p<*0.001.*

*95% confidence intervals presented in brackets.*

*Omega total ($\omega_t$) internal consistency estimates presented in parentheses along diagonal.*

Fit indices for all mediation path models are reported in Table 2. The baseline full mediation model (Model 1) in which internalised HIV/AIDS-related stigma and body appreciation were specified as serial mediators linking self-esteem with ART adherence evidenced an unsatisfactory level of fit. Inclusion of a direct effect between self-esteem and ART adherence (Model 2) worsened model fit. Although model fit improved when a path was added to Model 1 linking self-esteem with body appreciation (Model 3), there remained an unsatisfactory level of fit to the data. A substantial improvement in model fit was found by adding a path to Model 1 linking internalised HIV/AIDS-related stigma with ART adherence (Model 4). All model fit indices indicated that Model 4 had a satisfactory level of fit to the data and was identified as the best fitting model. We proceeded with comparing all alternative models to Model 4.

We compared Model 4 to variations of Models 1 and 4 in which the order of self-esteem and ART adherence was inverted (Models 5 and 6), as well as an alternative to Model 4 in which the order of internalised HIV/AIDS-related stigma and body appreciation was interchanged (Model 7). Fit indices revealed that each of these alternative models had an unsatisfactory level of fit to the data. Akaike information criterion values also supported a superior level of fit for Model 4, which we retained as the best fitting model.

The standardised path coefficients for Model 4 are reported in Figure 1. There was a small negative association between self-esteem and internalised HIV/AIDS-related stigma. Internalised HIV/AIDS-related stigma evidenced a moderate negative association with body appreciation and a small negative association with ART adherence. A moderate positive association was found between body appreciation and ART adherence. Taken together, the findings are consistent with indirect-only mediation and suggest that self-esteem is indirectly associated with ART adherence via internalised HIV/AIDS-related stigma and body appreciation.

## Discussion

The purpose of this study was to examine associations between self-esteem, internalised HIV/AIDS-related stigma, body appreciation, and ART adherence in a sample of YPLHIV in order to acquire further insight into potential pathways for promoting ART adherence among young people. As hypothesised, the results revealed that self-esteem and ART adherence were linked indirectly via a sequential path of internalised HIV/AIDS-related stigma followed by body appreciation. There was also evidence of a one-step indirect effect linking self-esteem with ART adherence via internalised HIV/AIDS-related stigma, although such an effect did not emerge for body appreciation. Overall, the results offer support for internalised HIV/AIDS-related stigma and body appreciation as mechanisms underlying the association between self-esteem and ART adherence.

The finding that self-esteem is not directly associated with ART adherence is consistent with a broader body of research on the mechanisms (e.g. counselling, social support) underlying the association between self-esteem and medication adherence.[56] Evidence of a two-step indirect effect suggests that the sequential path of internalised HIV/AIDS-related stigma and then body appreciation represents one mechanism that accounts for the association between self-esteem and ART adherence. This sequential pathway can be interpreted in the context of existing conceptual models that highlight the role of low self-esteem in precipitating health risk behaviour. Conceptualising self-esteem within the framework of identity theory[57], both the initial distress of receiving an HIV-seropositive diagnosis and the ongoing experience of living with HIV have the potential to disrupt the process of self-verification. Disturbances to self-verification can lead to a depletion of self-esteem, thereby reducing the capacity of self-esteem to operate as a buffer against stressors that emerge.[21] Among YPLHIV, low self-esteem may increase vulnerability to internalised HIV/AIDS-related stigma that arises out of the self-deprecating consequences of evaluating the self against stigmatising social norms (e.g. promiscuity) about HIV.[22] Internalisation of HIV/AIDS-related stigma is associated with fear of developing identifiable markers of the disease (e.g. lipodystrophy), which can heighten a person's sensitivity to bodily changes and increase bodily shame.[29] Previous research has found that the psychological sequalae of body image disturbance (e.g. depression) can have negative implications for ART adherence[8], suggesting that similar mechanisms may be involved in the association between body appreciation and ART adherence.

**Table 2:** Fit indices for serial mediation models ($N=76$)

| | Model fit indices | | | | | Comparative fit index |
|---|---|---|---|---|---|---|
| | $\chi^2$ (*df*) | CFI | TLI | RMSEA [90% CI] | SRMR | AIC |
| Model 1[a] | 7.30 (3) | 0.889 | 0.777 | 0.137 [0.000, 0.287] | 0.058 | 996.73 |
| Model 2[b] | 6.72* (2) | 0.878 | 0.634 | 0.176 [0.024, 0.348] | 0.060 | 998.65 |
| Model 3[c] | 5.38 (2) | 0.913 | 0.738 | 0.149 [0.000, 0.325] | 0.050 | 997.58 |
| **Model 4[d]** | **2.48 (2)** | **0.988** | **0.963** | **0.056 [0.000, 0.290]** | **0.030** | **994.71** |
| Model 5[e] | 16.13* (3) | 0.660 | 0.320 | 0.240 [0.121, 0.378] | 0.090 | 1008.15 |
| Model 6[f] | 13.21* (2) | 0.710 | 0.130 | 0.272 [0.132, 0.439] | 0.080 | 1007.94 |
| Model 7[g] | 13.21* (2) | 0.710 | 0.130 | 0.272 [0.132, 0.439] | 0.080 | 1003.49 |

*CFI, comparative fit index; TLF, Tucker Lewis index; RMSEA, root mean square error of approximation; SRMR, standardised root mean square residual; AIC, Akaike information criterion*

[a]*Self-esteem → Internalised HIV/AIDS-related stigma → Body appreciation → Antiretroviral therapy adherence.*

[b]*Model 1 with direct path added: Self-esteem → Antiretroviral therapy adherence.*

[c]*Model 1 with direct path added: Self-esteem → Body appreciation.*

[d]*Model 1 with direct path added: Internalised HIV/AIDS-related stigma → Antiretroviral therapy adherence.*

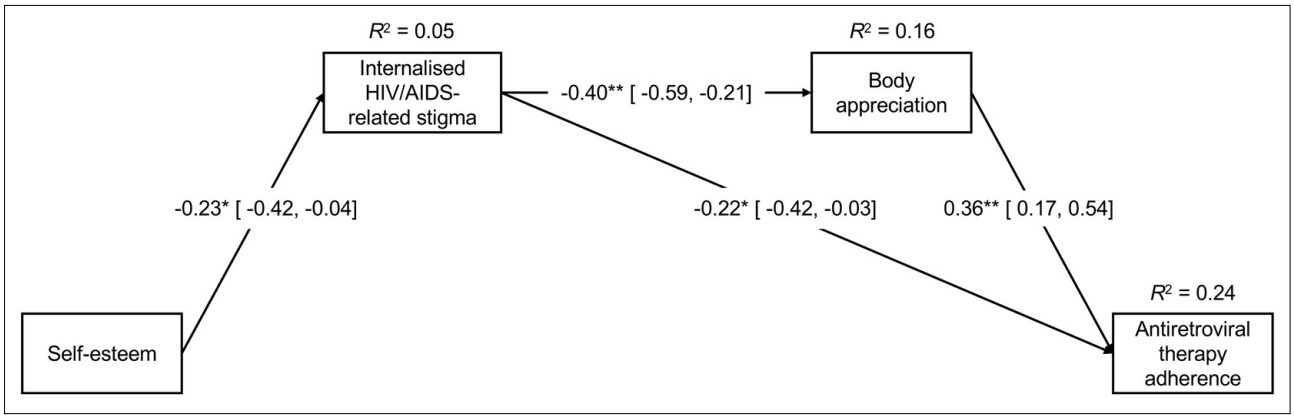
[e]*Alternative to Model 1 in which the ordering of self-esteem and antiretroviral therapy adherence is reversed.*

[f]*Alternative to Model 4 in which the ordering of self-esteem and antiretroviral therapy adherence is reversed.*

[g]*Alternative to Model 4 in which the ordering of Internalised HIV/AIDS-related stigma and body appreciation is reversed.*

*$p<0.05$

*Note: Entries in boldface reflect best fitting model.*




*Note. Standardised coefficients and 95% confidence intervals reported; *p <0.05, **p<0.001*

**Figure 1:** Final serial mediation model of associations between self-esteem, internalised HIV/AIDS-related stigma, body appreciation, and antiretroviral therapy adherence.

Taken together, the two-step indirect effect found in this study suggests that YPLHIV who receive support to reduce or protect against internalised HIV/AIDS-related stigma may benefit by experiencing higher levels of body appreciation, which in turn could lead to better ART adherence.

A unique contribution of this study to the evidence on ART adherence among YPLHIV is the inclusion of body appreciation as a sequential mechanism through which self-esteem and ART adherence are linked. A previous study highlighted the importance of identifying mechanisms by which internalised HIV/AIDS-related stigma is associated with ART adherence to enhance the efficacy of treatment approaches for improving treatment adherence.[58] Based on the findings of this study, other intrapersonal (e.g. depression) and interpersonal (e.g. social support) factors that could affect ART adherence among YPLHIV should be targeted alongside body appreciation issues to promote ART adherence.

The finding of a two-step indirect effect via internalised HIV/AIDS-related stigma and then body appreciation was supplemented by a one-step indirect effect via internalised HIV/AIDS-related stigma, which aligns with prior research involving self-esteem, internalised stigma, and medication adherence.[23] Considering the findings of this study did not support a one-step mediating effect for body appreciation, internalised HIV/AIDS-related stigma appears to play a key role in linking self-esteem and ART adherence. Should the availability of resources for promoting ART adherence be limited by infrastructure or economic constraints, treatment approaches targeting ART adherence might consider prioritising internalised HIV/AIDS-related stigma over body appreciation.

## Practical implications

Although longitudinal evidence is needed to build on this study, the findings suggest that intervention initiatives targeting internalised HIV/AIDS-related stigma and body appreciation may offer viable pathways for improving ART adherence. Several standalone interventions have successfully addressed internalised HIV/AIDS-related stigma and body appreciation in isolation. Cognitive-behavioural therapy approaches have been effective at reducing internalised HIV/AIDS-related stigma by challenging maladaptive patterns of thinking and building internal self-regulatory resources (e.g. coping skills) to manage stigma.[59] Fitness training interventions[60] have been successful at improving body image perceptions (including body appreciation and satisfaction with body functionality) and generalised self-efficacy by encouraging people to focus more on the functionality of their body and less on their appearance. Many of these standalone approaches could be integrated into interventions that incorporate other key components (e.g. social support) to provide a multi-pronged and more comprehensive approach to targeting ART adherence decisions and behaviours among YPLHIV.

## Limitations and future research directions

The findings of this study offer additional insight into the mechanisms underlying the association between self-esteem and ART adherence, but there are several limitations to acknowledge. First, the cross-sectional nature of the data limits our ability to establish causal inferences and draw definitive conclusions about directionality. Although several models were compared to identify the best fitting model from the alternatives, longitudinal studies are needed to establish the directionality and causal pathways linking the variables included in this study. Second, the sample comprised a small, relatively homogeneous group of black African YPLHIV who were recruited from a single source located in a populous urban setting. Coupled with the purposive convenience sampling approach that was used to recruit participants, caution should be applied in generalising the findings to other populations of PLHIV. Additional research is needed to determine the replicability of the findings in specific populations, particularly among vulnerable young key populations. Third, effect estimates ranged from small to moderate[61], and it is possible that the findings of this study are confounded by the omission of relevant variables that might exert a stronger mediating effect on the association between self-esteem and ART adherence (e.g. self-efficacy, autonomous motivation). Future studies might consider integrating additional variables into modelling procedures to explore the relative strength of alternative indirect mechanisms that link self-esteem with ART adherence. Along similar lines, further research is needed to identify other factors that may affect a person's experiences of bodily changes, attitude towards living with HIV, and ART adherence behaviour (e.g. socio-demographic factors, mode of HIV infection, length of time since HIV-seropositive diagnosis). Fourth, the findings should also be considered alongside the small sample of participants in this study. HIV remains a highly stigmatised health condition in South Africa and is a sensitive topic for many young people who often have concerns about the possible consequences of participating in research on HIV, including discrimination and having their HIV status further exposed.[62] Only 63% of the YPLHIV who were approached consented to participate in this study, with many offering the aforementioned concerns as reasons for declining to participate.

## Conclusion

In summary, the findings of the current cross-sectional study provide preliminary evidence of both one-step (through internalised HIV/AIDS-related stigma) and two-step (through internalised HIV/AIDS-related stigma and then body appreciation) indirect effects linking self-esteem and ART adherence among YPLHIV. Notwithstanding the need to determine whether the results reported herein replicate in other key populations by employing more sophisticated methodological approaches, the findings raise the intriguing possibility of promoting ART adherence in young people through targeted intervention initiatives that

seek to resolve internalised HIV/AIDS-related stigma and improve body appreciation.

## Acknowledgements

We acknowledge the young participants in this study, without whom the study would not have been possible. We also thank the staff of the HIV youth centre who assisted us in gaining access to the respondents.

## Competing interests

We have no competing interests to declare.

## Authors' contributions

P.N.: Conceptualisation; methodology; data collection; writing – the initial draft. K.G.: Conceptualisation; student supervision; critical revisions; R.G.C.: Conceptualisation; methodology; data analysis; data curation; student supervision; writing – revisions.

## References

1. UNAIDS. 90–90–90: An ambitious treatment target to help end the AIDS epidemic [document on the Internet]. c2017 [cited 2020 Feb 10]. Available from: https://www.unaids.org/sites/default/files/sub_landing/90-90-90_en.pdf
2. Schaecher KL. The importance of treatment adherence in HIV. Am J Manag Care. 2013;19:S231–S237.
3. Nachega JB, Marconi VC, Van Zyl GU, Gardner EM, Preiser W, Hong SY, et al. HIV treatment adherence, drug resistance, virologic failure: Evolving concepts. Infect Disord Drug Targets. 2011;11:167–174. https://doi.org/10.2174/187152611795589663
4. Adejumo OA, Malee KM, Ryscavage P, Hunter SJ, Taiwo BO. Contemporary issues on the epidemiology and antiretroviral adherence of HIV-infected adolescents in sub-Saharan Africa: A narrative review. J Int AIDS Soc. 2015;18(1):1–19. https://doi.org/10.7448/ias.18.1.20049
5. Govender K, Masebo WG, Nyamaruze P, Cowden RG, Schunter BT, Bains A. HIV prevention in adolescents and young people in the Eastern and Southern African Region: A review of key challenges impeding actions for an effective response. Open AIDS J. 2018;12:53–67. https://doi.org/10.2174/1874613601812010053
6. Cowden RG, Tucker LA, Govender K. Conceptual pathways to HIV risk in Eastern and Southern Africa: An integrative perspective on the development of young people in contexts of social-structural vulnerability. In: Govender K, Poku NK, editors. Preventing HIV among young people in Eastern and Southern Africa. New York: Routledge; 2020. p. 31–47. https://doi.org/10.4324/9780429462818-4
7. Mutumba M, Bauermeister JA, Musiime V, Byaruhanga J, Francis K, Snow RC, et al. Psychosocial challenges and strategies for coping with HIV among adolescents in Uganda: A qualitative study. AIDS Patient Care STDS. 2015;29:86–94. https://doi.org/10.1089/apc.2014.0222
8. Blashill AJ, Goshe BM, Robbins GK, Mayer KH, Safren SA. Body image disturbance and health behaviors among sexual minority men living with HIV. Health Psychol. 2014; 33:677-80. https://doi.org/10.1037/hea0000081
9. Shubber Z, Mills EJ, Nachega JB, Vreeman R, Freitas M, Bock P, et al. Patient-reported barriers to adherence to antiretroviral therapy: A systematic review and meta-analysis. PLoS Med. 2016;13, e1002183. https://doi.org/10.1371/journal.pmed.1002183
10. Henriksen IO, Ranøyen I, Indredavik MS, Stenseng F. The role of self-esteem in the development of psychiatric problems: A three-year prospective study in a clinical sample of adolescents. Child Adolesc Psychiatry Ment Health. 2017;11:68. https://doi.org/10.1186/s13034-017-0207-y
11. Zhang L, Li X, Lin Z, Jacques-Tiura AJ, Xu J, Zhou Y, et al. Side effects, adherence self-efficacy, and adherence to antiretroviral treatment: A mediation analysis in a Chinese sample. AIDS Care. 2016;28:919–926. https://doi.org/10.1080/09540121.2015.1124984
12. Rosenberg M. Society and adolescent child. Princeton, NJ: Princeton University Press; 1965.
13. Mann MM, Hosman CM, Schaalma HP, De Vries NK. Self-esteem in a broad-spectrum approach for mental health promotion. Health Educ Res. 2004;19:357–372. https://doi.org/10.1093/her/cyg041
14. Hagen R, Havnen A, Hjemdal O, Kennair LE, Ryum T, Solem S. Protective and vulnerability factors in self-esteem: The role of metacognitions, brooding, and resilience. Front Psychol. 2020;11:1447. https://doi.org/10.3389/fpsyg.2020.01447
15. Enejoh V, Pharr J, Mavegam BO, Olutola A, Karick H, Ezeanolue EE. Impact of self esteem on risky sexual behaviors among Nigerian adolescents. AIDS Care. 2016;28:672–676. https://doi.org/10.1080/09540121.2015.1120853
16. Sileo KM, Kizito W, Wanyenze RK, Chemusto H, Reed E, Stockman JK, et al. Substance use and its effect on antiretroviral treatment adherence among male fisherfolk living with HIV/AIDS in Uganda. PLoS ONE. 2019;14(6), e0216892. https://doi.org/10.1371/journal.pone.0216892
17. Zogg JB, Woods SP, Sauceda JA, Wiebe JS, Simoni JM. The role of prospective memory in medication adherence: A review of an emerging literature. J Behav Med. 2012;35:47–62. https://doi.org/10.1007/s10865-011-9341-9
18. Poquette AJ, Moore DJ, Gouaux B, Morgan EE, Grant I, Woods SP, et al. Prospective memory and antiretroviral medication non-adherence in HIV: An analysis of ongoing task delay length using the memory for intentions screening test. J Int Neuropsychol Soc. 2013;19(2):155–161. https://doi.org/10.1017/S1355617712001051
19. Tsai AC, Bangsberg DR, Bwana M, Haberer JE, Frongillo EA, Muzoora C, et al. How does antiretroviral treatment attenuate the stigma of HIV? Evidence from a cohort study in rural Uganda. AIDS Behav. 2013;17:2725–2731. https://doi.org/10.1007/s10461-013-0503-3
20. Quinn DM. Issue introduction: Identity concealment: Multilevel predictors, moderators, and consequences. J Soc Issues. 2017;73:230–239. https://doi.org/10.1111/josi.12213
21. Crocker J, Quinn DM. Social stigma and the self: Meanings, situations, and self-esteem. In: Heatherton TF, Kleck RE, Hebl MR, Hull JG, editors. The social psychology of stigma. New York: Guilford Press; 2000. p. 153–183.
22. Brown JD, Dutton KA, Cook KE. From the top down: Self-esteem and self-evaluation. Cogn Emot. 2001;15:615–31. https://doi.org/10.1080/02699930126063
23. Visser M, Sipsma H. The experience of HIV-related stigma in South Africa. In: Liamputtong P, editor. Stigma, discrimination and living with HIV/AIDS stigma, discrimination and living with HIV/AIDS. Dordrecht: Springer; 2013. p. 205–227. https://doi.org/10.1007/978-94-007-6324-1_12
24. Goudge J, Ngoma B, Manderson L, Schneider H. Stigma, identity and resistance among people living with HIV in South Africa. SAHARA J. 2009;6(3):94–104. https://doi.org/10.1080/17290376.2009.9724937
25. Bernier A, Benmoussa A, Hilali MK, Henry E, Otis J, Loukid M, et al. Self-esteem and HIV infection in Morocco: Associated factors among people living with HIV – Results from a community-based study. Community Ment Health J. 2019;55(8):1402–1410. https://doi.org/10.1007/s10597-019-00394-9
26. Pantelic M, Shenderovich Y, Cluver L, Boyes M. Predictors of internalised HIV-related stigma: A systematic review of studies in sub-Saharan Africa. Health Psychol. Rev. 2015;9(4):469–490. https://doi.org/10.1080/17437199.2014.996243
27. Varas-Díaz N, Toro-Alfonso J, Serrano-García I. My body, my stigma: Body interpretations in a sample of people living with HIV/AIDS in Puerto Rico. Qual Rep. 2005;10:122–142. https://doi.org/10.1177/1049732304272059
28. Zeng C, Li L, Hong YA, Zhang H, Babbitt AW, Liu C, et al. A structural equation model of perceived and internalized stigma, depression, and suicidal status among people living with HIV/AIDS. BMC Public Health. 2018;18:138. https://doi.org/10.1186/s12889-018-5053-1
29. Palmer AK, Duncan KC, Ayalew B, Zhang W, Tzemis D, Lima V, et al. "The way I see it": The effect of stigma and depression on self-perceived body image among HIV-positive individuals on treatment in British Columbia, Canada. AIDS Care. 2011;23:1456–1466. https://doi.org/10.1080/09540121.2011.565021
30. Chapman E. Conceptualisation of the body for people living with HIV: Issues of touch and contamination. Sociol Health Illn. 2000;22:840–857. https://doi.org/10.1111/1467-9566.00233
31. Tylka TL, Wood-Barcalow NL. What is and what is not positive body image? Conceptual foundations and construct definition. Body Image. 2015;14:118–129. https://doi.org/10.1016/j.bodyim.2015.04.001

32. Nyamaruze P, Govender K. "I like the way I am, but I feel like I could get a little bit bigger": Perceptions of body image among adolescents and youth living with HIV in Durban, South Africa. PLoS ONE. 2020;15, e0227583. https://doi.org/10.1371/journal.pone.0227583

33. Bankole KO, Bakare MO, Edet BE, Igwe MN, Ewa AU, Bankole IA, et al. Psychological complications associated with HIV/AIDS infection among children in South-South Nigeria, sub-Saharan Africa. Cogent Med. 2017;4(1):1372869. https://doi.org/10.1080/2331205X.2017.1372869

34. Lamb KM, Nogg KA, Safren SA, Blashill AJ. Mechanism of change in cognitive behavioral therapy for body image and self-care on ART adherence among sexual minority men living with HIV. AIDS Behav. 2018;22:2711–2717. https://doi.org/10.1007/s10461-018-2143-0

35. Gillen MM. Associations between positive body image and indicators of men's and women's mental and physical health. Body Image. 2015;13:67–74. https://doi.org/10.1016/j.bodyim.2015.01.002

36. Meland E, Haugland S, Breidablik HJ. Body image and perceived health in adolescence. Health Educ Res. 2007;22:342–350. https://doi.org/10.1093/her/cyl085

37. Ramjee G, Sartorius B, Morris N, Wand H, Reddy T, Yssel JD, et al. A decade of sustained geographic spread of HIV infections among women in Durban, South Africa. BMC Infect Dis. 2019;19(1):1–9. https://doi.org/10.1186/s12879-019-4080-6

38. Westaway MS, Jordaan ER, Tsai J. Investigating the psychometric properties of the Rosenberg self-esteem scale for South African residents of greater Pretoria. Eval Health Prof. 2015;38:181–199. https://doi.org/10.1177/0163278713504214

39. Supple AJ, Su J, Plunkett SW, Peterson GW, Bush KR. Factor structure of the Rosenberg Self-Esteem Scale. J Cross-Cult Psychol. 2013;44:748–764. https://doi.org/10.1177/0022022112468942

40. Ugoji FN. Determinants of risky sexual behaviours among secondary school students in Delta State Nigeria. Int J Adolesc Youth. 2014;19:408–418. https://doi.org/10.1080/02673843.2012.751040

41. Kalichman SC, Simbayi LC, Cloete A, Mthembu PP, Mkhonta RN, Ginindza T. Measuring AIDS stigmas in people living with HIV/AIDS: The Internalized AIDS Related Stigma Scale. AIDS Care. 2009;21:87–93. https://doi.org/10.1080/09540120802032627

42. Kalichman SC, Simbayi LC, Jooste S, Toefy Y, Cain D, Cherry C, et al. Development of a brief scale to measure AIDS-related stigma in South Africa. AIDS Behav. 2005;9:135–143. https://doi.org/10.1007/s10461-005-3895-x

43. Tylka TL, Wood-Barcalow NL. The Body Appreciation Scale-2: Item refinement and psychometric evaluation. Body Image. 2015;12:53–67. https://doi.org/10.1016/j.bodyim.2014.09.006

44. Halliwell E, Jarman H, Tylka T, Slater A. Adapting the Body Appreciation Scale-2 for children: A psychometric analysis of the BAS-2C. Body Image. 2017;21:97–102. https://doi.org/10.1016/j.bodyim.2017.03.005

45. Swami V, Ng SK, Barron D. Translation and psychometric evaluation of a standard Chinese version of the Body Appreciation Scale-2. Body Image. 2016;18:23–26. https://doi.org/10.1016/j.bodyim.2016.04.005

46. Morisky DE, Ang A, Krousel-Wood M, Ward HJ. Predictive validity of a medication adherence measure in an outpatient setting. J Clin Hypertens. 2008;10:348–354. https://doi.org/10.1111/j.1751-7176.2008.07572.x

47. Saal W, Kagee A. The applicability of the Theory of Planned Behaviour in predicting adherence to ART among a South African sample. J Health Psychol. 2012;17:362–370. https://doi.org/10.1177/1359105311416875

48. De las Cuevas C, Peñate W. Psychometric properties of the eight-item Morisky Medication Adherence Scale (MMAS-8) in a psychiatric outpatient setting. Int J Clin Health Psychol. 2015;15(2):121–129. https://doi.org/10.1016/j.ijchp.2014.11.003

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

50. Tabachnick BG, Fidell LS. Using multivariate statistics. 6th ed. Boston, MA: Pearson; 2013.

51. Mueller RO, Hancock GR. Best practices in structural equation modeling. In: Osborne JW, editor. Best practices in quantitative methods. Thousand Oaks, CA: SAGE; 2008. p. 488–510. https://doi.org/10.4135/9781412995627.d38

52. Worthington RL, Whittaker TA. Scale development research: A content analysis and recommendations for best practices. Couns Psychol. 2006;34:806–838. https://doi.org/10.1177/0011000006288127

53. Hopwood CJ, Donnellan MB. How should the internal structure of personality inventories be evaluated? Pers Soc Psychol Rev. 2010;14:332–346. https://doi.org/10.1177/1088868310361240

54. Weston R, Gore Jr PA. A brief guide to structural equation modeling. Couns Psychol. 2006;34:719–751. https://doi.org/10.1177/0011000006286345

55. Burnham KP, Anderson DR, Huyvaert KP. AIC model selection and multimodel inference in behavioral ecology: Some background, observations, and comparisons. Behav Ecol Sociobiol. 2011;65:23–35. https://doi.org/10.1007/s00265-010-1029-6

56. Hussain S, Malik A, Hussain Z. A randomized controlled intervention trial: Effect of counselling on treatment adherence and self-esteem of women patients receiving tuberculosis treatment. Open Med. 2016;3:27–33. https://doi.org/10.2174/1874220301603010027

57. Cast AD, Burke PJ. A theory of self-esteem. Soc Forces. 2002;80:1041–1068. https://doi.org/10.1353/sof.2002.0003

58. Helms CB, Turan JM, Atkins G, Kempf MC, Clay OJ, Raper JL, et al. Interpersonal mechanisms contributing to the association between HIV-related internalized stigma and medication adherence. AIDS Behav. 2017;21:238–247. https://doi.org/10.1007/s10461-016-1320-2

59. Tshabalala J, Visser M. Developing a cognitive behavioural therapy model to assist women to deal with HIV and stigma. S Afr J Psychol. 2011;41:17–28. https://doi.org/10.1177/008124631104100103

60. Bassett RL, Ginis KA. Risky business: The effects of an individualized health information intervention on health risk perceptions and leisure time physical activity among people with spinal cord injury. Disabil Health J. 2011;4:165–176. https://doi.org/10.1016/j.dhjo.2010.12.001

61. Cohen J. Statistical power analysis. Curr Dir Psychol Sci. 1992;1(3):98–101. https://doi.org/10.1111/1467-8721.ep10768783

62. DiClemente RJ, Sales JM, Borek N. Barriers to adolescents' participation in HIV biomedical prevention research. J Acquir Immune Defic Syndr. 2010;54:S12–S17. https://doi.org/10.1097/QAI.0b013e3181e1e2c0

**AUTHORS:**
Michael C. Everett[1]
Marina C. Elliott[2,3]
David Gaynor[4]
Austin C. Hill[1,5]
Samar M. Syeda[6]
Jesse Casana[1]
Bernhard Zipfel[7]
Jeremy M. DeSilva[1,7]
Nathaniel J. Dominy[1]

**AFFILIATIONS:**
[1]Department of Anthropology, Dartmouth College, Hanover, New Hampshire, USA
[2]Department of Archaeology, Simon Fraser University, Burnaby, British Columbia, Canada
[3]Centre for the Exploration of the Deep Human Journey, University of the Witwatersrand, Johannesburg, South Africa
[4]Mammal Research Institute, University of Pretoria, Pretoria, South Africa
[5]Department of Anthropology, University of Pennsylvania, Philadelphia, Pennsylvania, USA
[6]Skeletal Biology Research Centre, School of Anthropology and Conservation, University of Kent, Canterbury, UK
[7]Evolutionary Studies Institute, University of the Witwatersrand, Johannesburg, South Africa

**CORRESPONDENCE TO:**
Michael Everett

**EMAIL:**
Michael.C.Everett.19@dartmouth.edu

**DATES:**
**Received:** 13 Mar. 2021
**Revised:** 28 July 2021
**Accepted:** 02 Aug. 2021
**Published:** 29 Nov. 2021

**HOW TO CITE:**
Everett MC, Elliott MC, Gaynor D, Hill AC, Syeda SM, Casana J, et al. Mechanical loading of primate fingers on vertical rock surfaces. S Afr J Sci. 2021;117(11/12), Art. #10409. https://doi.org/10.17159/sajs.2021/10409

**ARTICLE INCLUDES:**
☒ Peer review
☒ Supplementary material

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

**EDITORS:**
Margaret Avery
Jemma Finch

**KEYWORDS:**
human evolution, locomotion, climbing, suspension, bone curvature

**FUNDING:**
Claire Garber Goodman Fund, Department of Anthropology, Dartmouth College; Stamps Scholars Program, Dartmouth College

# Mechanical loading of primate fingers on vertical rock surfaces

Mechanical loading of finger bones (phalanges) can induce angular curvature, which benefits arboreal primates by dissipating forces and economising the recruitment of muscles during climbing. The recent discovery of extremely curved phalanges in a hominin, *Homo naledi*, is puzzling, for it suggests life in an arboreal milieu, or, alternatively, habitual climbing on vertical rock surfaces. The importance of climbing rock walls is attested by several populations of baboons, one of which uses a 7-m vertical surface to enter and exit Dronkvlei Cave, De Hoop Nature Reserve, South Africa. This rock surface is an attractive model for estimating the probability of extreme mechanical loading on the phalanges of rock-climbing primates. Here we use three-dimensional photogrammetry to show that 82–91% of the climbable surface would generate high forces on the flexor tendon pulley system and severely load the phalanges of baboons and *H. naledi*. If such proportions are representative of vertical rock surfaces elsewhere, it may be sufficient to induce stress-mitigating curvature in the phalanges of primates.

**Significance:**

- We present the first three-dimensional photogrammetric analysis of a vertical rock surface climbed by a non-human primate, the chacma baboon (*Papio ursinus*).
- Our results show that a large proportion of a vertical rock wall would compel crimp and slope hand positions during climbing – grips that could explain the extraordinary phalangeal curvature expressed by a Middle Pleistocene hominin, *Homo naledi.*

## Introduction

Bone is a dynamic tissue, and repeated mechanical loading can induce changes to its density and angular curvature. For example, loading stresses from vertical climbing and suspensory locomotion can increase the curvature of primate finger bones (phalanges) during growth and development.[1-3] Greater curvature is advantageous to arboreal primates because it dissipates forces and dampens recruitment of muscles during finger flexion, thus economising energetic costs and simultaneously lowering the risk of falling. Phalangeal curvature is therefore widely viewed as a measure of arboreality among primates – it is a classic form-functional trait that informs our interpretations of behaviour and ecology in the fossil record[1-7] (but see Wallace et al.[8] for a counterexample).

Upsetting this orthodoxy is the hand of *Homo naledi*, recovered from Rising Star Cave, South Africa, in 2013–14 and described in 2015.[9] The degree of phalangeal curvature is astounding, with an included angle ($\theta$) that far exceeds that of modern humans (Figure 1). Such curvature would normally and unequivocally suggest an arboreal milieu[3], but there are at least two reasons to doubt such an inference for *H. naledi*. First, the phalanges are relatively short, resulting in humanlike hand proportions.[9] No arboreal primate has phalanges that are both short and curved; it is an aberration that confounds conventional interpretation. Second, existing specimens of *H. naledi* are dated to the Middle Pleistocene between 335 kya and 241 kya.[12] Nearby faunal assemblages of comparable age – those of Gladysvale[13] and Lincoln Cave, Sterkfontein[14] – favour habitat conditions that resemble those of today, i.e. a mix of acacia woodland and grassland, neither of which would incentivise a strongly arboreal ecology.[15]

So, what was *H. naledi* doing with its hands? Voisin et al.[16] examined the shoulder girdle of *H. naledi* and reported morphological traits associated with vertical climbing and suspension. But given the inferred habitat conditions, they discounted arboreal activities and argued instead for 'movement across and climbing on rocky walls'[16(p.2)]. Voisin et al.'s hypothesis is intriguing given that rock climbing puts high forces on finger flexor tendons, especially during the 'crimp' position (Figure 2).[17-20] This position puts extreme stress on the flexor tendon pulley system, a series of five annular (A1–A5) and three cruciate ligaments (C1–C3) that resist bowstringing of the tendons. In other words, the pulley system holds the tendons close to the bone, effectively converting linear force into torque that produces flexion at the metacarpophalangeal and interphalangeal joints (Supplementary figure 1).

Ruptures of A2[17] and other severe pulley injuries are common among human rock climbers[21], and a testament to the mechanical loading of phalanges. Indeed, frequent rock climbing is known to cause geometric and cortical thickening of the phalanges[22] and it may induce greater curvature, at least hypothetically. Tan et al.[18] calculated the benefits of greater phalangeal curvature for reducing tendon and pulley stresses of fingers in the crimp position; however, affirming increased curvature among rock climbers is difficult because it is challenging to measure angles from radiographs of living subjects. Another model system is needed.

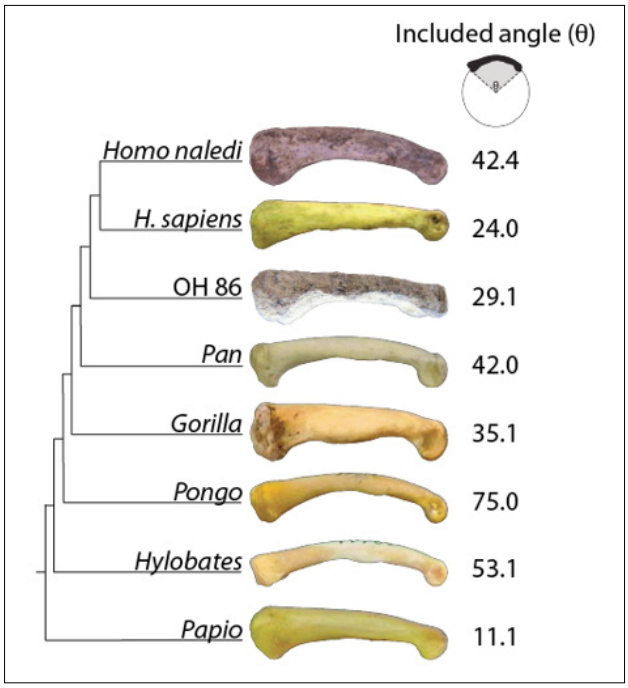



**Figure 1:** Cladogram of baboons (*Papio*), gibbons (*Hylobates*), orangutans (*Pongo*), gorillas (*Gorilla*), chimpanzees (*Pan*), an indeterminate hominin (OH 86), and genus *Homo*. The branches are tipped with representative photographs of manual proximal phalanges in lateral view, together with mean degrees of curvature of the proximal phalanx, as measured by computing the included angle ($\theta$). Data sources for $\theta$: Stern et al.[10] (humans), Jungers et al.[6] (non-human primates), and Domínguez-Rodrigo et al.[11] (OH 86). The $\theta$ value for *H. naledi* was calculated by one of us (S.M.S.) as the mean of digits II (42.71), III (46.49), IV (48.44), and V (31.88) of hand 1, the right hand of an adult (see Kivell et al.[9] for further details).

Some populations of baboons and chimpanzees enter cave systems to avoid predators and/or regulate body temperature.[23-34] Some caves are only accessible by climbing sheer cliffs.[28] For example, Marais[23] described a group of chacma baboons (*Papio ursinus*) and their daily ascent to a cliffside cave opening ≈150 m high. His vivid account speaks to the mechanical loading of fingers: 'they had to go more than a mile hanging only by their fingers to the [cliff] ledge; their hind feet against the smooth surface'[23(p.63)]. In some cases, baboons must negotiate vertical rock surfaces within a cave. For example, at Misgrot Cave in Thabazimbi, South Africa, baboons face a perilous 17-m vertical descent from ground level (Figure 3)[32], whereas those entering Dronkvlei Cave, De Hoop Nature Reserve, South Africa, descend 7 m to access the interior (Figure 4). Baboons use both caves regularly as overnight sleeping sites; Dronkvlei is used on 28% of nights.[30]

### *Study aims and design*

Dronkvlei Cave is an appealing model system for exploring the topography of a vertical rock surface and its potential to load primate fingers. Here we ask a basic question: what proportion of a vertical rock wall would, when climbed, force the distal phalanges of *P. ursinus* (and, hypothetically, *H. naledi)* into a crimp or slope grip position? It is a thought experiment that leaves formal comparative measures of phalangeal curvature in *P. ursinus* as a priority for future research. To answer our question, we used three-dimensional (3D) photogrammetry to detect and quantify concavities in the surface that range from a minimum graspable depth (determined as the length of the third distal phalanx) to a minimum crimpable depth (determined as the sum length of the third intermediate and distal phalanx) – i.e. depths of 11–25 mm for *P. ursinus* and 14–37 mm for *H. naledi* (Table 1).

## Methods

### *Data acquisition*

Photogrammetry is a low-cost method for extracting accurate measurements from photographs. It is well suited for producing 3D models of irregular surfaces, such as cave walls.[36] To obtain photographs of the rock surface climbed by baboons, we used metal irrigation piping to construct a temporary T-shaped scaffold. The cross pipe was mounted

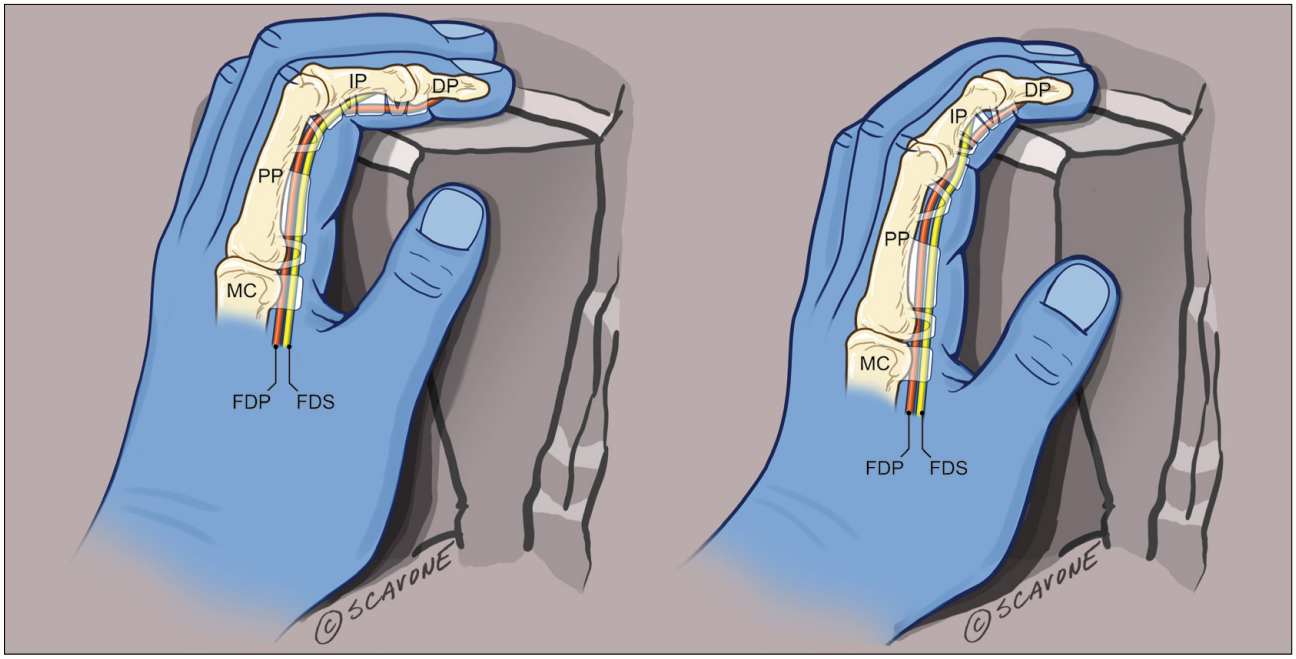
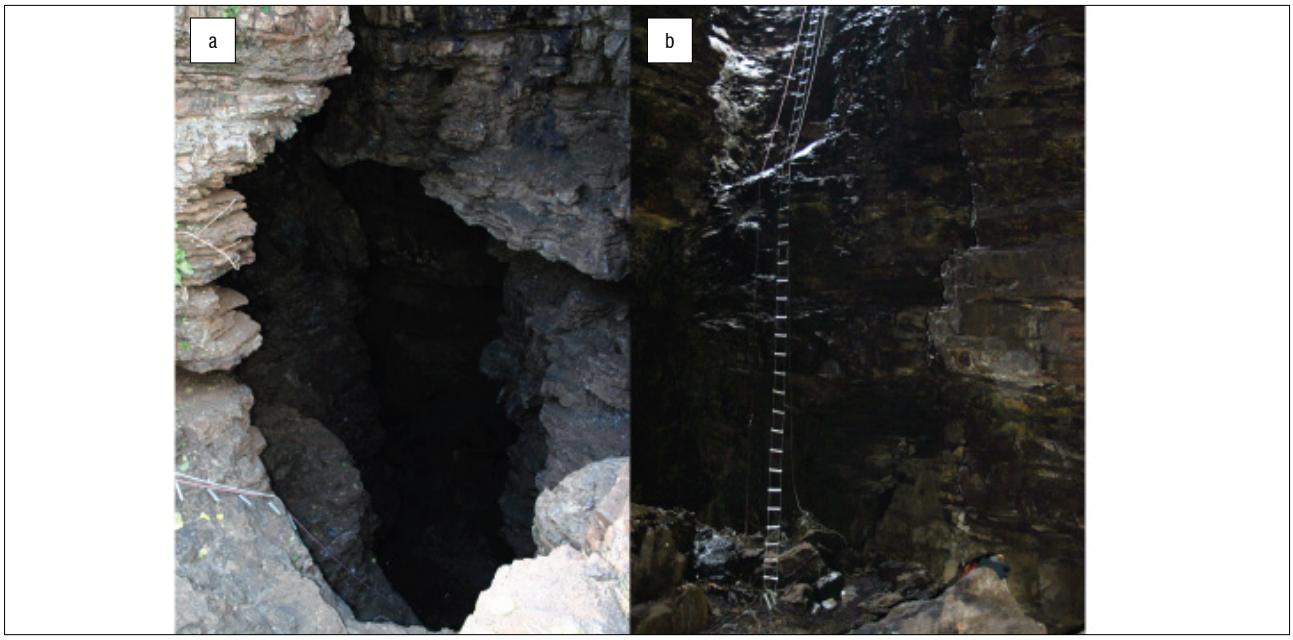



*Illustration: ©William Scavone, reproduced with permission.*

*Anatomical labels: MC, metacarpal; PP, proximal phalanx; IP, intermediate phalanx; DP, distal phalanx; FDS, flexor digitorum superficialis; FDP, flexor digitorum profundus*

**Figure 2:** Common handholds on rock surfaces and the underlying flexor tendon pulley system. Up to 90% of climbers use the crimp grip (*left*), where the proximal interphalangeal (PIP) joints are flexed from 90° to 100° and the distal interphalangeal (DIP) joints are either fully extended or hyperextended.[17] The second most common grip is the slope grip or open hand grip (*right*), which is distinguished by extended or slightly flexed PIP joints and flexion (50° to 70°) of the DIP joints.[17]




*Photographs: Jason L. Heaton, reproduced with permission.*

**Figure 3:** Misgrot Cave in Thabazimbi, South Africa. Chacma baboons (a) enter the cave at ground level before negotiating (b) a steep vertical descent of 17 m.

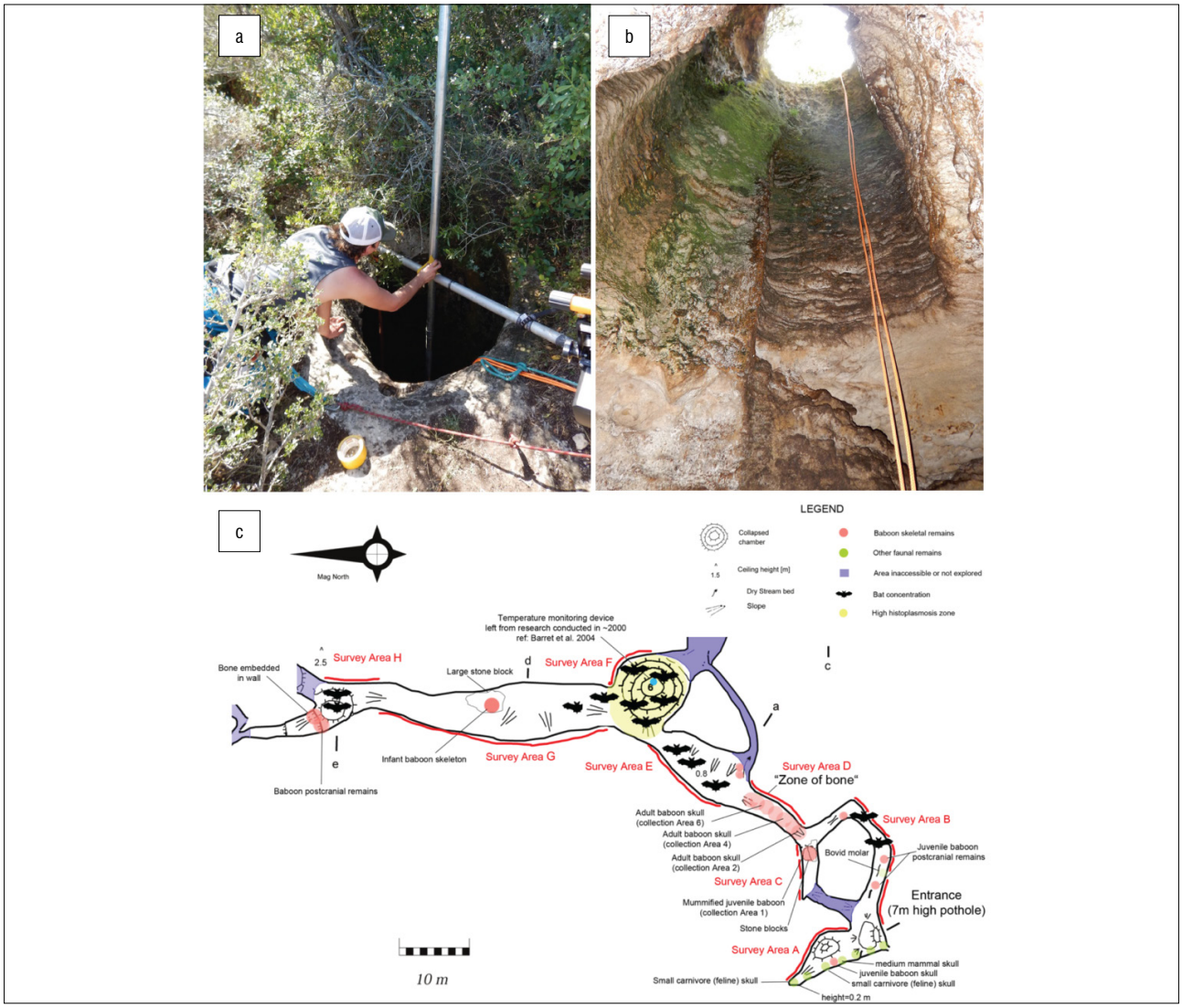



**Figure 4:** Dronkvlei Cave, De Hoop Nature Reserve, South Africa. Chacma baboons (a) enter the cave through a 1-m opening at ground level before negotiating (b) a vertical descent of 7 m. The preferred surface of the baboons is clearly differentiated by a dark patina. This surface area is the focus of our analysis. In addition, we produced (c) a map of the interior, noting the presence of skeletal remains. Some remains are now accessioned at the University of the Witwatersrand (repository prefix U.W., site designation 116, see Zipfel and Berger[35] for cataloguing details).

at the cave opening and fixed to a vertical pipe that extended to the shaft floor (Figure 4a), where it was weighted to minimise lateral excursions. We hitched a digital camera to this pipe so that photography occurred from a fixed central axis, and we mounted ground control points to the rock surface. We abseiled into the cave to operate the camera, aiming for 80% overlap between serial photos in the horizontal and vertical planes.

**Table 1:** Lengths (in mm) of the third intermediate and distal phalanx of the hand. Most measures of *Papio ursinus* were obtained from specimens in the School of Anatomical Sciences, University of the Witwatersrand (see Supplementary table 1 for catalogue numbers and corresponding measurements).

| Species (*N* individuals) | Intermediate phalanx mean ± 1 SD (range) | Distal phalanx mean ± 1 SD (range) | Source |
|---|---|---|---|
| *Papio ursinus* (19) | 14.4 ± 1.7 (12.2–19.7) | 10.8 ± 1.3 (9.2–13.8) | This study |
| *Homo naledi* (2) | 22.2 (21.8–22.6) | 14.4 | Kivell et al.[9], Kivell, unpublished |

### *Photogrammetry*

We processed the image set ($n$=354) with Agisoft Metashape Pro (formerly Photoscan Pro), a common package used in archaeological research.[37-39] Due to logistical constraints in the field, we did not record spatial data for each ground control point, so the model was scaled to real-world dimensions by using the targets as scale bars, and manually oriented. Manual orientation is suboptimal for any calculations that rely on the slope of the wall, but it should not affect the local depth calculations described below. After alignment, dense cloud construction, and texture application, we trimmed the 3D model to the area used for climbing (Figure 4b) and exported it as an orthophoto (Figure 5a) and digital elevation model (DEM) with an arbitrary local coordinate system. We brought the resulting geotiffs, both with sub-millimetre resolution, into ArcGIS and SAGA GIS for further processing and visualisation.

### *Depth processing*

In ArcGIS, the DEM raster treats the surface of the rock wall as vertical elevation data. To calculate the local relative depth of each raster cell in the rock wall, we used focal statistics with a small neighborhood, and maximum elevation as the statistics type, to create a raster of the local maximum height of the wall, and then smoothed that raster again with focal statistics over a wider neighborhood, creating a smoothed idealised model of the maximum height of the local surface of the cave. We subtracted the original DEM from this idealised surface to find the local difference between each cell and the average nearest 'maximum' value. Finally, we queried this raster of local depths for cells in two ranges (11–25 mm for *P. ursinus* and 14–37 mm for *H. naledi*) and used this value to calculate the total area of each range.

## Results and discussion

The entrance to Dronkvlei Cave is a useful model for exploring potential stresses on the flexor tendon pulley system and phalanges of primates. We focused our analysis on the surface area climbed by baboons – representing 5.98 m$^2$ (Figure 5a) – and we show that crimp- and slope-inducing handholds are distributed relatively evenly across a nearly vertical plane (Figure 5b). However, the total area of such holds differed between the two species.

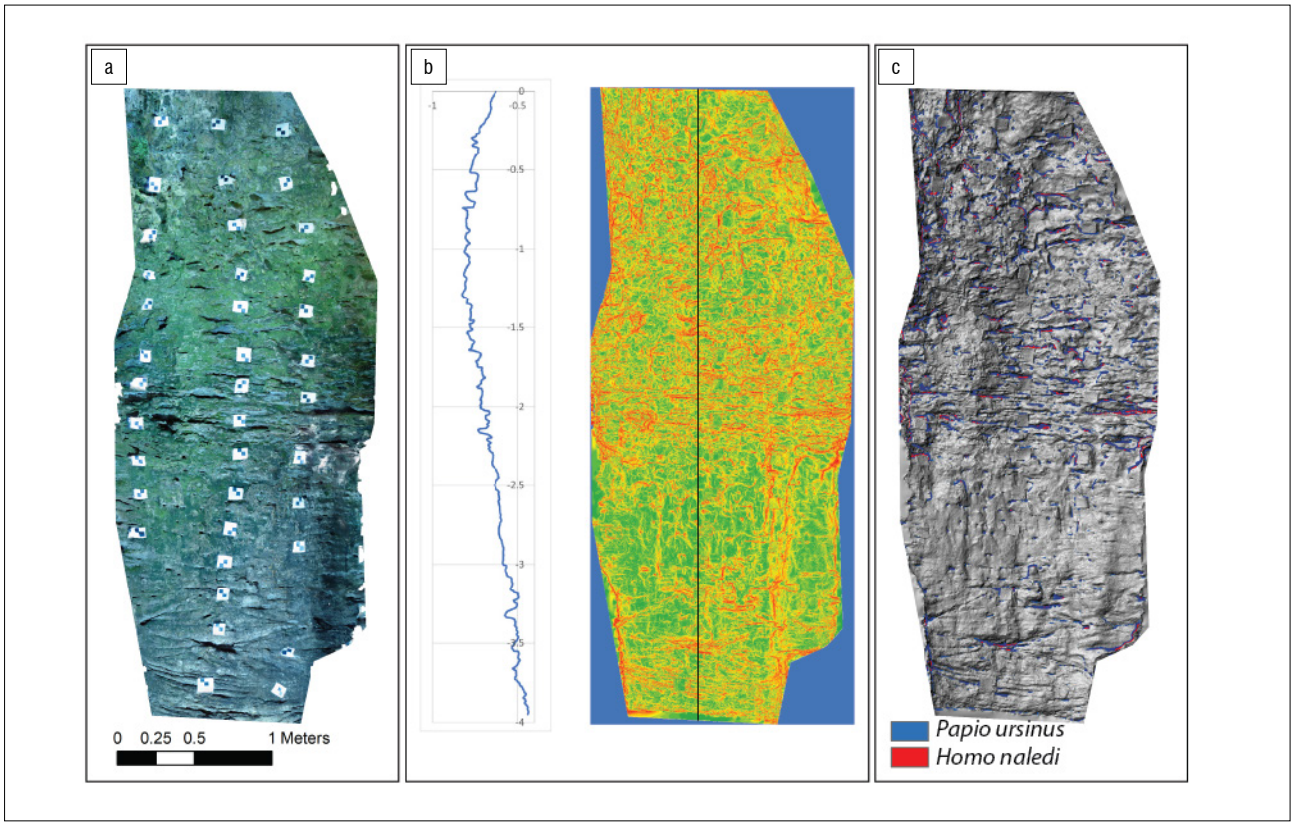



**Figure 5:** Section of vertical rock wall at the entrance to Dronkvlei Cave, South Africa. It is a surface that baboons climb regularly (Figure 4b) and the focus of our analysis. (a) Composite orthophoto and ground control points. (b) Digital elevation model to visualise the rugosity and verticality of the surface (the cross-sectional profile on the left corresponds to the black line). Red surfaces slope outward, whereas green surfaces are relatively flat. (c) Hillshaded visualisation; the colours correspond to handholds that are likely to compel weight-bearing finger flexion by chacma baboons (*Papio ursinus*) and *Homo naledi*. The 3D mesh and raw photo are available in MorphoSource (https://www.morphosource.org/projects/000348582).

For *P. ursinus*, we calculated an area of 0.494 $m^2$ (2 514 603 cells of 0.44 mm x 0.44 mm), and for *H. naledi* an area of 0.353 $m^2$ (1 795 106 cells of 0.44 mm x 0.44 mm), with an overlap of 0.282 $m^2$ (Figure 3c). Thus, 6–8% of the total surface area, when gripped, is expected to generate high forces on the flexor tendon pulley system and severely load the phalanges of both primate species. Such an estimate is conservative, however, as a large proportion of the rock surface is smooth and unclimbable. When we subtract surface depths of <11 mm and <14 mm from the total, the total graspable surface area is reduced to 0.602 $m^2$ and 0.390 $m^2$, respectively, meaning that as much as 82% (*P. ursinus*) or 91% (*H. naledi*) of usable handholds would produce forces that would favour greater phalangeal curvature.

The significance of our finding is a matter of speculation – even at 8% of the total surface area, the rock wall of Dronkvlei Cave is a non-trivial source of phalangeal loading when extrapolated over a lifetime. The average life expectancy of a female baboon that survives to adulthood is 12.1 years in Amboseli, Kenya, and 19.7 years in Gombe, Tanzania (with record longevities of 27 years at both sites).[40] If we use these parameters to estimate an average life expectancy of ≈16 years for *P. ursinus*, and given that animals climb 12 m to enter and exit Dronkvlei Cave on 28% of days[30], an individual would encounter ≈1.6 km of crimp- and slope-grip-inducing surfaces over its lifetime [(16 years × 365 days × 0.28) × (12 m × 0.08 m)]. It is a crude estimate of mechanical 'loading' on the flexor tendon pulley system and distal phalanges, but it invites an analysis of phalangeal curvature in this population.

More instructive perhaps are the mummified/skeletal remains of baboons in Misgrot Cave.[34] These individuals climb greater distances than those at Dronkvlei (Figure 3b), but the frequency of overnight use is unknown. Still, each climbing bout carries great risk. Falling from heights as high as 17 m is likely to cause severe morbidity or outright mortality[15], although the rugosity of the dolomitic surface may differ from the limestone surface of Dronkvlei Cave in crucial ways, affecting risk. Misgrot Cave is a promising site for studying the natural history of why, and the mechanics of how, baboons climb vertical rock surfaces, as well as the effects on their finger bones. Another promising study system lies in the lava tubes of Mount Suswa, Kenya, the site of an oft-cited analysis by Simons[24], who described the skeletal remains of baboons (*P. anubis*) and leopards, sometimes commingled, at the base of vertical shafts. Remains of baboons continue to accumulate there, serving as a tourist attraction.

So, did *H. naledi* climb vertical rock surfaces regularly? Answering this question is beyond reach at present, but their bodies, alive or dead, negotiated at least two steep walls of dolomite to arrive in the Dinaledi Chamber of Rising Star Cave: an entry point through the roof of Postbox Chamber and a 12-m fissure ('The Chute').[12] Our results suggest that climbing these surfaces would have compelled the use of crimp or slope grips to some extent, and Tan et al.'s[18] model of flexor tendon pulley forces suggests that *H. naledi* would have enjoyed a 23% reduction in phalangeal forces compared to modern humans (Supplementary figure 2). Such findings indicate that high phalangeal curvature – whether mechanically induced or genetically mediated[8] – would have conferred advantages to *H. naledi* during rock climbing.

A limitation of our model system approach is that *H. naledi* (weighing ~40 kg[41]) was heavier than *P. ursinus* is now and probably longer-lived, factors that would load its fingers to a greater extent. Yet the pedal phalanges of *H. naledi* express the same curvature as those of *Papio*, which is '[possibly] indicative of elevated pedal grasping ability'[42(p.4)]. Thus, both species are equipped to grip vertical rock surfaces with their feet – a prudent means of distributing body mass and mitigating forces on the fingers during climbing.

We conclude by acknowledging that rock-climbing baboons are an imperfect model for interpreting the curious anatomy of *H. naledi*, but we would argue that detailed studies of their functional anatomy and climbing kinematics, together with systematic measures of the rock surfaces used, are promising areas of future research.

## Acknowledgements

We are grateful to Sergio Almécija, Louise Barrett, Lee-Anne Benjamin, Jackson Dolphin, Elli Goudzwaard, Caitlyn Hauke, Jason Heaton, Peter Henzi, Kudakwashe Jakata, Ashley Kehoe, Tracy Kivell, Scott McGraw, Maureen O'Leary, Randall Susman, and Mary Wilde for generous practical assistance. For access to comparative baboon remains, our thanks go to Brendon Billings, School of Anatomical Sciences, University of the Witwatersrand. This research was approved by CapeNature (permit no. CN44-30-184) and funded by the Claire Garber Goodman Fund, Department of Anthropology, Dartmouth College (to M.C.E) and the Stamps Scholars Program, Dartmouth College (to M.C.E.).

## Competing interests

We have no competing interests to declare.

## Authors' contributions

M.C.Everett: Conceptualisation, methodology, data collection, data curation, data analysis, writing – the initial draft, writing – revisions, project leadership. M.C.Elliott: Methodology, data collection, writing – revisions. D.G.: Methodology, data collection, writing – revisions. A.C.H.: Methodology, data analysis, writing – the initial draft. S.M.S.: Data collection. J.C.: Methodology, writing – the initial draft. B.Z.: Data curation, data collection, writing – revisions. J.M.D.: Conceptualisation, student supervision, writing – the initial draft, writing – revisions. N.J.D.: Conceptualisation, student supervision, writing – the initial draft, writing – revisions.

## References

1. Richmond BG. Biomechanics of phalangeal curvature. J Hum Evol. 2007;53:678–690. http://doi.org/10.1016/j.jhevol.2007.05.011
2. Congdon KA. Interspecific and ontogenetic variation in proximal pedal phalangeal curvature of great apes (*Gorilla gorilla*, *Pan troglodytes*, and *Pongo pygmaeus*). Int J Primatol. 2012;33:418–427. http://doi.org/10.1007/s10764-012-9590-7
3. Patel BA, Maiolino SA. Morphological diversity in the digital rays of primate hands. In: Kivell TL, Lemelin P, Richmond BG, Schmitt D, editors. The evolution of the primate hand: Anatomical, developmental, functional, and paleontological evidence. New York: Springer; 2016. p. 55–100.
4. Susman RL. Comparative and functional morphology of hominoid fingers. Am J Phys Anthropol. 1979;50:215–236. http://doi.org/10.1002/ajpa.1330500211
5. Jungers WL, Godfrey LR, Simons EL, Chatrath PS. Phalangeal curvature and positional behavior in extinct sloth lemurs (Primates, Palaeopropithecidae). Proc Natl Acad Sci USA. 1997;94:11998–12001. http://doi.org/10.1073/pnas.94.22.11998
6. Jungers WL, Godfrey LR, Simons EL, Wunderlich RE, Richmond BG, Chatrath PS. Ecomorphology and behavior of giant extinct lemurs from Madagascar. In: Plavcan JM, Kay RF, Jungers WL, Van Schaik CP, editors. Reconstructing behavior in the primate fossil record. Boston, MA: Springer; 2002. p. 371–411. http://doi.org/10.1007/978-1-4615-1343-8_10
7. Prang TC, Ramirez K, Grabowski M, Williams SA. *Ardipithecus* hand provides evidence that humans and chimpanzees evolved from an ancestor with suspensory adaptations. Sci Adv. 2021;7:eabf2474. http://doi.org/10.1126/sciadv.abf2474
8. Wallace IJ, Burgess ML, Patel BA. Phalangeal curvature in a chimpanzee raised like a human: Implications for inferring arboreality in fossil hominins. Proc Natl Acad Sci USA. 2020;117:11223–11225. http://doi.org/10.1073/pnas.2004371117
9. Kivell TL, Deane AS, Tocheri MW, Orr CM, Schmid P, Hawks J, et al. The hand of *Homo naledi*. Nat Commun. 2015;6:8431. http://doi.org/10.1038/ncomms9431
10. Stern JT Jr, Jungers WL, Susman RL. Quantifying phalangeal curvature: An empirical comparison of alternative methods. Am J Phys Anthropol. 1995;97:1–10. http://doi.org/10.1002/ajpa.1330970102

11. Domínguez-Rodrigo M, Pickering TR, Almécija S, Heaton JL, Baquedano E, Mabulla A, et al. Earliest modern human-like hand bone from a new >1.84-million-year-old site at Olduvai in Tanzania. Nat Commun. 2015;6:7987. http://doi.org/10.1038/ncomms8987

12. Robbins JL, Dirks PHGM, Roberts EM, Kramers JD, Makhubela TV, Hilbert-Wolf HL, et al. Providing context to the *Homo naledi* fossils: Constraints from flowstones on the age of sediment deposits in rising Star Cave, South Africa. Chem Geol. 2021;567, Art. #120108. http://doi.org/10.1016/j.chemgeo.2021.120108

13. Lacruz, RS, Brink JS, Hancox PJ, Skinner AR, Herries A, Schmid P, et al. Palaeontology and geological context of a Middle Pleistocene faunal assemblage from the Gladysvale Cave, South Africa. Palaeontol Afr. 2002;38:99–114. http://hdl.handle.net/10539/16351

14. Reynolds SC, Clarke RJ, Kuman KA. The view from the Lincoln Cave: Mid- to late Pleistocene fossil deposits from Sterkfontein hominid site, South Africa. J Hum Evol. 2007;53:260–271. http://doi.org/10.1016/j.jhevol.2007.02.004

15. Kraft TS, Venkataraman VV, Dominy NJ. A natural history of human tree climbing. J Hum Evol. 2014;71:105–118. http://doi.org/10.1016/j.jhevol.2014.02.002

16. Voisin JL, Feuerriegel EM, Churchill SE, Berger LR. The *Homo naledi* shoulder girdle: An adaptation to boulder climbing. L'Anthropologie 2020;124:102783. http://doi.org/10.1016/j.anthro.2020.102783

17. Schweizer A. Biomechanical properties of the crimp grip position in rock climbers. J Biomech. 2001;34:217–223. http://doi.org/10.1016/S0021-9290(00)00184-6

18. Tan MA, Fuss FK, Niegl G. Stress distribution at the finger pulleys during sport climbing. In: Fuss FK, Subic A, Ujihashi S, editors. The impact of technology on sport II. London: CRC Press; 2007. p. 663–669. https://doi.org/10.1201/9781439828427.ch97

19. Schöffl I, Oppelt K, Jüngert J, Schweizer A, Neuhuber W, Schöffl V. The influence of the crimp and slope grip position on the finger pulley system. J Biomech. 2009;42:2183–2187. http://doi.org/10.1016/j.jbiomech.2009.04.049

20. Amca AM, Vigouroux L, Aritan S, Berton E. Effect of hold depth and grip technique on maximal finger forces in rock climbing. J Sports Sci. 2012;30:669–677. http://doi.org/10.1080/02640414.2012.658845

21. Schöffl V, Schöffl I, Frank L, Küpper T, Simon M, Lutter C. Tendon injuries in the hands in rock climbers: epidemiology, anatomy, biomechanics and treatment – an update. Muscles Ligaments Tendons J. 2020;10:233–243. http://doi.org/10.32098/mltj.02.2020.08

22. Morrison AB, Schöffl VR. Physiological responses to rock climbing in young climbers. Br J Sports Med. 2007;41:852–861. https://doi.org/10.1136/bjsm.2007.034827

23. Marais EN. My friends the baboons. London: Methuen; 1939.

24. Simons JW. The presence of leopard and a study of the food debris in the leopard lairs of the Mount Suswa caves, Kenya. Bull Cave Exploration Group E Afr. 1966;1:51–69.

25. Gow CE. Habitual sheltering in an extensive cave system by baboons near Bredasdorp, South Africa. S Afr J Sci. 1973;69:182.

26. Wells L. Baboons sheltering in caves. S Afr J Sci. 1973;69:279.

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

28. Hamilton WJ. Baboon sleeping site preferences and relationships to primate grouping patterns. Am J Primatol. 1982;3:41–53. http://doi.org/10.1002/ajp.1350030104

29. McGrew WC, McKee JK, Tutin CEG. Primates in caves: Two new reports of *Papio* spp. J Hum Evol. 2003;44:521–526. http://doi.org/10.1016/S0047-2484(03)00042-3

30. Barrett L, Gaynor D, Rendall D, Mitchell D, Henzi SP. Habitual cave use and thermoregulation in chacma baboons (*Papio hamadryas ursinus*). J Hum Evol. 2004;46:215–222. http://doi.org/10.1016/j.jhevol.2003.11.005

31. Hodgins G, Brook GA, Marais E. Bomb-spike dating of a mummified baboon in Ludwig Cave, Namibia. Int J Speleol. 2007;36:31–38. http://scholarcommons.usf.edu/ijs/vol36/iss1/3

32. Egeland CP, Domínguez-Rodrigo M, Pickering TR, Menter CG, Heaton JL. Hominin skeletal part abundances and claims of deliberate disposal of corpses in the Middle Pleistocene. Proc Natl Acad Sci USA. 2018;115:4601–4606. http://doi.org/10.1073/pnas.1718678115

33. Boyer Ontl K, Pruetz JD. Mothers frequent caves: Lactation affects chimpanzee (*Pan troglodytes verus*) cave use in southeastern Senegal. Int J Primatol. 2020;41:916–935. http://doi.org/10.1007/s10764-020-00165-4

34. Nel C, Bradfield J, Lombard M, Val A. Taphonomic study of a modern baboon sleeping site at Misgrot, South Africa: Implications for large-bodied primate taphonomy in karstic deposits. J Paleolith Archaeol. 2021;4:4. http://doi.org/10.1007/s41982-021-00080-x

35. Zipfel B, Berger LR. New Cenozoic fossil-bearing site abbreviations for collections of the University of the Witwatersrand. Palaeontol Afr. 2009;44:77–81.

36. González-Aguilera D, Muñoz-Nieto A, Gómez-Lahoz J, Herrero-Pascual J, Gutierrez-Alonso G. 3D digital surveying and modelling of cave geometry: Application to Paleolithic rock art. Sensors. 2009;9:1108–1127. http://doi.org/10.3390/s90201108

37. Doneus M, Verhoeven G, Fera M, Briese Ch, Kucera M, Neubauer W. From deposit to point cloud – a study of low-cost computer vision approaches for the straightforward documentation of archaeological excavations. Geoinformatics FCE CTU. 2011;6:81–88. http://doi.org/10.14311/gi.6.11

38. De Reu J, Plets G, Verhoeven G, De Smedt P, Bats M, Cherretté B, et al. Towards a three-dimensional cost-effective registration of the archaeological heritage. J Archaeol Sci. 2013;40:1108–1121. http://doi.org/10.1016/j.jas.2012.08.040

39. Olson BR, Placchetti RA, Quartermaine J, Killebrew AE. The Tel Akko Total Archaeology Project (Akko, Israel): Assessing the suitability of multi-scale 3D field recording in archaeology. J Field Archaeol. 2013;38:244–262. http://doi.org/10.1179/0093469013Z.00000000056

40. Bronikowski AM, Alberts SC, Altmann J, Packer C, Carey KD, Tatar M. The aging baboon: Comparative demography in a non-human primate. Proc Natl Acad Sci USA. 2002;99:9591–9595. http://doi.org/10.1073/pnas.142675599

41. Garvin HM, Elliott MC, Delezene LK, Hawks J, Churchill SE, Berger LR, et al. Body size, brain size, and sexual dimorphism in *Homo naledi* from the Dinaledi Chamber. J Hum Evol. 2017;111:119–138. http://doi.org/10.1016/j.jhevol.2017.06.010

42. Harcourt-Smith WEH, Throckmorton Z, Congdon KA, Zipfel B, Deane AS, Drapeau MSM, et al. The foot of *Homo naledi*. Nat Commun. 2015;6:8432. http://doi.org/10.1038/ncomms9432

**AUTHORS:**
Martina Meincken[1]
Gerhard Roux[2]
Thomas Niesler[3]

**AFFILIATIONS:**
[1]Department of Forest and Wood Science, Stellenbosch University, Stellenbosch, South Africa
[2]Department of Music, Stellenbosch University, Stellenbosch, South Africa
[3]Department of Electrical and Electronic Engineering, Stellenbosch University, Stellenbosch, South Africa

**CORRESPONDENCE TO:**
Martina Meincken

**EMAIL:**
mmein@sun.ac.za

**DATES:**
**Received:** 19 May 2021
**Revised:** 02 Aug. 2021
**Accepted:** 16 Sep. 2021
**Published:** 29 Nov. 2021

**HOW TO CITE:**
Meincken M, Roux G, Niesler T. An African violin – The feasibility of using indigenous wood from southern Africa as tonewood. S Afr J Sci. 2021;117(11/12), Art. #11175. https://doi.org/10.17159/sajs.2021/11175

**ARTICLE INCLUDES:**
☒ Peer review
☐ Supplementary material

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

**EDITOR:**
Michael Inggs

**KEYWORDS:**
wood for soundboards, spectral analysis, exotic wood, musical instruments, string instruments

**FUNDING:**
None

# An African violin – The feasibility of using indigenous wood from southern Africa as tonewood

The wood used to make musical instruments needs to have particular properties. Depending on its function, such as a soundboard for string instruments or the body of a wind instrument, different properties are desirable to obtain the best musical quality. Several different classification schemes exist that correlate physical and mechanical properties of wood to define desirable ranges for tonewoods, and to allow suitable wood species to be chosen. The physical and mechanical properties of various wood species indigenous to southern Africa were characterised and then assessed in terms of their suitability for violin construction using these classification schemes. The results of this analysis show that the most suitable of the wood species assessed are yellowwood and sapele. These were subsequently used by a professional luthier to build an 'African' violin. The sound quality of this instrument was determined subjectively through performances to an audience and more objectively via spectral analysis of audio recordings. This analysis shows clear differences in the relative magnitude of the harmonics between the violin made from indigenous wood and an instrument made with conventional wood species. Despite the differences, yellowwood and sapele were found to be suitable tonewoods, resulting in an instrument with a unique sound.

**Significance:**

- Good quality violins are always made from spruce and maple wood, which have to be imported to South Africa – often at high cost. However, the growth conditions of most southern African wood species should make them suitable tonewoods. We showed that several species are suitable to be used as tonewoods and that the sound produced with such a violin is – although somewhat different – of high quality.

## Introduction

Tonewoods are wood species that possess certain desirable properties that make them suitable for the construction of musical instruments, such as woodwind or string instruments. Any good tonewood should be radially cut, have an even-grain structure, be defect free and be dimensionally stable, i.e. not shrink or swell noticeably with environmental changes.[1] Soundboards – woods that should resonate in order to amplify the oscillation – of string instruments should be lightweight, but stiff enough to withstand the tension of the strings and propagate sound well in the grain direction. The speed of sound along the wood grain should be as high as possible and substantially higher parallel to the grain than perpendicular to the grain.[2] Pronounced density differences in the wood lead to a scattering of the soundwave; therefore, significant differences between heart- and sapwood, or early- and latewood, are not desirable. This should make (sub)tropical wood species good candidates for tonewood, because the lack of pronounced growth seasons results in wood with barely visible year rings and few density variations.

In this paper, we report on mostly indigenous southern African wood species that are suitable for string instruments of the violin family.

High-quality violins are made of the same wood species worldwide: spruce as the soundboard (top plate) and maple as the frameboard (back).[3] While guitar makers seem to be more adventurous in using alternative wood species, violin makers tend to use only these traditional wood species that often have to be imported at high cost. Good-quality wood is typically from colder regions, like Canada or the European Alps, where the trees grow more slowly, which results in a uniform wood structure with fewer density fluctuations.

Several classification schemes have been published that group wood species into clusters that indicate potential uses for the species and their physical and mechanical properties. Soundboards for a violin, for example, need to be good sound transmitters with a high radiation ratio (the speed of sound parallel to the grain divided by the density). Frameboards, on the other hand, should have a high Young modulus to give support to the soundboard, while still radiating sound well. Action wood, such as the keys and hammers in a piano, needs to be strong to withstand repeated use and woodwind instruments are typically made from a dense, finely structured wood that reflects the soundwaves well.

Yoshikawa[4] determined that all suitable soundboards for string instruments follow a regression line orthogonal to the regression line of frameboards when the transmission parameter cR is plotted against the antivibration parameter ρ/c (c: speed of sound, ρ: density, R: radiation ratio).

Wegst[5] developed a wider classification scheme for all musical uses, including string instruments, action wood, bows and woodwind. Based on wood properties – such as the speed of sound, elastic modulus and density – Wegst identified areas in these graphs that define suitable wood species for certain purposes. Soundboards, for example, need to have a high speed of sound with a reasonably low density (to minimise weight) and an acceptable Young modulus to provide strength.

Bucur[2] published an extensive review of the field detailing the required wood properties for musical instruments in his book *Acoustics of wood* where he discussed the physical and mechanical properties that make wood suitable as tonewood.

None of these classification schemes, however, includes southern African wood species. We characterised several wood species using all relevant mechanical and physical properties and aligned these with the classification schemes described above. The two most suitable species (for the soundboard and frameboard) were then chosen for construction of a full-sized violin.

Luthiers regard the radiation ratio[6] (speed of sound parallel to the grain divided by the density) as the most crucial property and its value should be as high as possible.[7] In good instruments, it can reach values of around 15.[8] However, equally important are dimensional stability, anatomical wood structure and wood elasticity, as they all affect the way the sound wave propagates through the wood.

Nevertheless, the quality of the instrument depends as much on the skill of the luthier as on the wood properties, which makes it challenging to assign sound quality to wood properties alone. Therefore, we sought the help of one of South Africa's best luthiers – Hannes Jacobs – to make the violin. This ensured that the quality of the instrument would be comparable to good-quality commercial instruments. Our instrument, named the 'African violin', was then compared to a violin made from traditional wood species by the same luthier, presumably with the same methodology.

The sound quality of the two violins was determined by computing audio frequency spectra of a set of single notes to analyse the location and amplitude of resonance peaks – or overtones – that are discernible. The aim was a quantitative rather than a qualitative comparison between the two instruments made by the same luthier, but from different wood species.

Although many publications have discussed the quality of violins, most use a descriptive terminology in which listeners use adjectives to describe the sound – to quantify differences between instruments.[9] Generally, it is accepted that a good sound consists of many harmonics at higher frequencies. Meinel[10] recorded audio signals for different violins of known quality – a Stradivarius and a cheaper instrument – and attempted to define the sound quality through the differences observed in the frequency spectra. His main findings were that a 'good' instrument has high amplitudes in the lower frequency range, low amplitudes around 1.5 kHz, which results in a 'nasal' sound, a strong peak in the region of 2–3 kHz and low amplitudes above 3 kHz. This study was taken further by Gabrielsson and Jansson[11], Jansson[12] and Dünnwald[13] who refined the definitions of desirable frequency bands somewhat, but largely agreed with the findings in Meinel's[10] study.

## Materials and methods

### Wood selection and characterisation

Wood was sourced from Stander Houtverkope in Knysna and Rare Woods in Cape Town. The following wood species indigenous to southern Africa were characterised: yellowwood (*Podocarpus latifolius*), candlewood (*Pterocelastrus tricuspidatus*), hardpear (*Olinia ventosa*), quar (*Psydrax obovata*), stinkwood (*Ocotea bullata*), and Cape beech (*Rapanea melanophloeos*). Furthermore sapele (*Entandrophragma cylindricum*) from western Africa and blackwood (*Acacia melanoxylon*) were included in the study. Although blackwood originates from Australia, it has been established in the natural forests of the southern Cape since the early 1900s and forms a valuable resource for the local woodworking industry.

For comparison, Italian spruce (*Picea abies*) and maple (*Acer platanoides*), tonewood typically used to make violins, were obtained from the luthier and also characterised. Hereafter, all species are referred to by their common names.

The moisture content (MC) and air-dry density ($\rho_{ad}$) were determined from 10 wood blocks per species that were air-dried and stored in a conditioning room at 20 °C and 65% relative humidity for 2 months before analysis. All results are reported as average values. The MC was determined as

$$MC = 100\% * (m_{ad}-m_0)/m_0$$

and the air-dry density was determined as

$$\rho_{ad} = m_{ad}/V_{ad}$$

where $m_{ad}$ = air-dry mass, $m_0$ = oven-dry mass and $V_{ad}$ = air-dry volume.

The acoustic properties were determined at three different positions on two conditioned boards per species with the dimensions of 40 x 15 x 2 $cm^3$ and are reported as average values.

The sound velocity ($c_s$) was determined with ultrasound waves that are transmitted and received with two probes over a known distance (l) in the time (t) as $c_s = l / t$. The MC is used as a correction value to determine the accurate speed of sound.

In these experiments, a Lucchi meter was used, which is the instrument typically used by luthiers for accurate measurement of smaller wooden pieces.

From the sound velocity and the wood density, several other properties can be calculated, such as the specific Young modulus, $E = c_s^2 * \rho$, and the radiation ratio, $R = c_s/\rho$ – properties which are used by the different classification schemes.

### Wood classification

The classification schemes reported by Wegst[5] and Yoshikawa[4] were used to identify the most suitable species for the construction of a violin. Wegst plotted various physical and mechanical properties against each other – such as the Young modulus, or the sound velocity, as a function of the density – and identified areas of desirable characteristics for different tonewood uses. Yoshikawa found that the plots of the transmission parameter $c_sR$ as a function of the antivibration parameter $\rho/c_s$ for the soundboards and frameboards of string instruments are linear and orthogonal.

The relevant properties of the wood species analysed in this study were fitted into these classification plots, and the species that best met the requirements were identified.

### The violin

A full-size violin was made from the most suitable species for soundboard and frameboard by an established luthier in South Africa, Hannes Jacobs. The sound produced by this instrument was then compared to that of a control instrument made by the same luthier, but with conventional wood species, namely spruce and maple.

### Sound analysis

Spectral analysis of the audio signals produced by both violins was performed on set of single notes to allow objective comparison of the produced harmonics. The violins were fitted with the same type of strings and played by the same person with the same bow.

Recordings were made in a studio at the Department of Music at Stellenbosch University with Sennheiser MKH 8020 omnidirectional microphones and a Sound Devices MixPre 10 T multitrack recorder at a sampling rate of 96 kHz with 24 bits per sample. The main microphone was placed 600 mm above the instrument and a control microphone was placed 600 mm in front of the musician on the same height as the violin.

The segmentation of the recordings into individual notes was accomplished using the open-source audio editor Audacity. Spectral analysis was performed using GNU Octave. The audio data of each note was divided into 16 384 sample (171 ms) frames with a 50% overlap. After application of a Hamming window, the power spectral density was computed for each frame using the fast Fourier transformation. Finally,

these power spectra were averaged over all analysed frames and the amplitudes normalised to the highest spectral peak.

Frequency spectra were plotted and compared using Origin.

## Results

### *Wood selection*

Table 1 shows the physical and mechanical properties of all analysed wood species. Rows indicating the two species commonly used for violin making, spruce and maple, are shaded. Both wood species were supplied by the luthier and are high quality tonewood. The spruce, which is used for the soundboard, has a low density and high Young modulus, giving it stability while not making it too heavy. The sound velocity along the grain (II) is high with a ratio of sound velocity parallel and perpendicular (⊥) to the grain of approximately 3. The resulting radiation ratio – the ratio of sound velocity along the grain to the density – is high with a value of R=13. The maple wood used for the frameboard (back plate) and the ribs has a higher density and Young modulus, giving stability to the instrument without adding too much weight. It has an acceptably high sound velocity and radiation ratio.

The southern African wood species are sorted according to their density. Candlewood and quar have very high densities, which makes them too heavy to be used for the instrument body, but they present a feasible option for accessories, such as the fingerboard or the bow. The high densities also lead to a low radiation ratio, making them unsuitable as tonewood.

Yellowwood and blackwood showed the highest radiation ratio among the analysed wood species – although significantly lower than the Italian spruce – with reasonably low densities, making them suitable to be used as soundboards. Blackwood also has a very high Young modulus, which together with the acceptable density makes it an interesting alternative as a soundboard. The lower radiation ratio of these wood species may, however, result in a duller sound, because fewer harmonics at high frequencies can be expected.

Stinkwood, hardpear, sapele and Cape beech all have higher densities than yellowwood and blackwood, acceptable Young moduli, and radiation ratios comparable to maple, which makes them suitable as frameboards.

The properties used by Yoshikawa[4] for the acoustical classification of various wood species are the transmission parameter, which he defined as c*R, and the antivibration parameter, defined as ρ/c. He plotted these properties for various wood species typically used for musical instruments and found that most of the wood species used for soundboards lie on a regression line that shows decreasing c*R with increasing ρ/c, while most of the wood species used for frameboards are located on a regression line that is nearly orthogonal to the regression line of the soundboards.

None of the values determined for the southern African species lie directly on these lines, but some are reasonably close. Based on this classification scheme, the most suitable species for the soundboard are blackwood, followed by yellowwood, stinkwood, Cape beech, hardpear and sapele, although the high density of the latter four might lead to an excessive weight of the instrument. The only feasible wood species for the frameboard based on this classification is sapele.

Wegst[5] plotted several relevant physical properties of various wood species against each other and found that the species known to be suitable for certain end uses – such as the violin front, the back plate and the violin bow – form well-defined clusters. We utilised the classification based on the Young modulus and density, as illustrated in Figure 2. In the original classification, the areas suitable for woodwind, piano action wood etc. were also included in the plot. For the purpose of this study, however, only the areas relevant to string instruments were considered.

Soundboards should have a low density and a high Young modulus, while frameboards tend to have a somewhat higher density and Young modulus, to add stability to the instrument. Based on this classification scheme, suitable wood species for the soundboard would be yellowwood and blackwood, while suitable species for the frameboard would be sapele, Cape beech, stinkwood and hardpear. Quar and candlewood would be ideal for violin bows, or accessories, such as the fingerboard.

Based on the two classification schemes, the physical properties and the visual appearance of the wood (i.e. the boards with the best radial cut and the straightest grain), it was decided to make a violin using yellowwood for the soundboard and sapele for the frameboard, neck and ribs. Figure 3 shows the first African violin. The wood was not stained, so the colour difference between the lightly coloured yellowwood and the reddish, dark sapele wood is maintained.

**Table 1:** Physical properties of the conventional tonewoods and southern African wood species

| Species | $r_{ad}$ (kg/m$^3$) | $c_{\perp}$ | $c_{II}$ | Ratio $c_{II}/_{\perp}$ | E (Gpa) | E/r | R | c/r | ρ/c | c*R ($10^5$) |
|---|---|---|---|---|---|---|---|---|---|---|
| Italian spruce | 423 | 1949 | 5588 | 2.87 | 13.21 | 0.03 | 13.2 | 13.21 | 0.08 | 7.38 |
| Maple | 626 | 1598 | 4820 | 3.02 | 14.54 | 0.02 | 7.7 | 7.70 | 0.13 | 3.71 |
| Quar | 963 | 1371 | 5042 | 3.68 | 24.49 | 0.03 | 5.24 | 5.24 | 0.19 | 2.64 |
| Candlewood | 949 | 1056 | 4759 | 4.51 | 21.49 | 0.02 | 5.01 | 5.01 | 0.20 | 2.39 |
| Hardpear | 740 | 2200 | 5560 | 2.53 | 22.57 | 0.03 | 7.62 | 7.62 | 0.13 | 4.23 |
| Stinkwood | 720 | 1700 | 5400 | 3.18 | 21.00 | 0.03 | 7.50 | 7.50 | 0.13 | 4.05 |
| Sapele | 680 | 1800 | 4700 | 2.61 | 15.02 | 0.02 | 6.91 | 6.91 | 0.14 | 3.25 |
| Cape Beech | 670 | 1200 | 5000 | 4.17 | 16.25 | 0.03 | 7.69 | 7.69 | 0.13 | 3.85 |
| Blackwood | 600 | 1900 | 5650 | 2.97 | 19.15 | 0.03 | 9.42 | 9.42 | 0.11 | 5.32 |
| Yellowwood | 505 | 2000 | 5000 | 2.50 | 12.63 | 0.03 | 9.90 | 9.90 | 0.10 | 4.95 |

*$r_{ad}$, air-dry density; $c_{\perp}$, sound velocity perpendicular to the grain; $c_{II}$, sound velocity parallel to the grain; E, Young modulus; R, radiation ratio*

*Note: Shading indicates the species commonly used for violin making.*

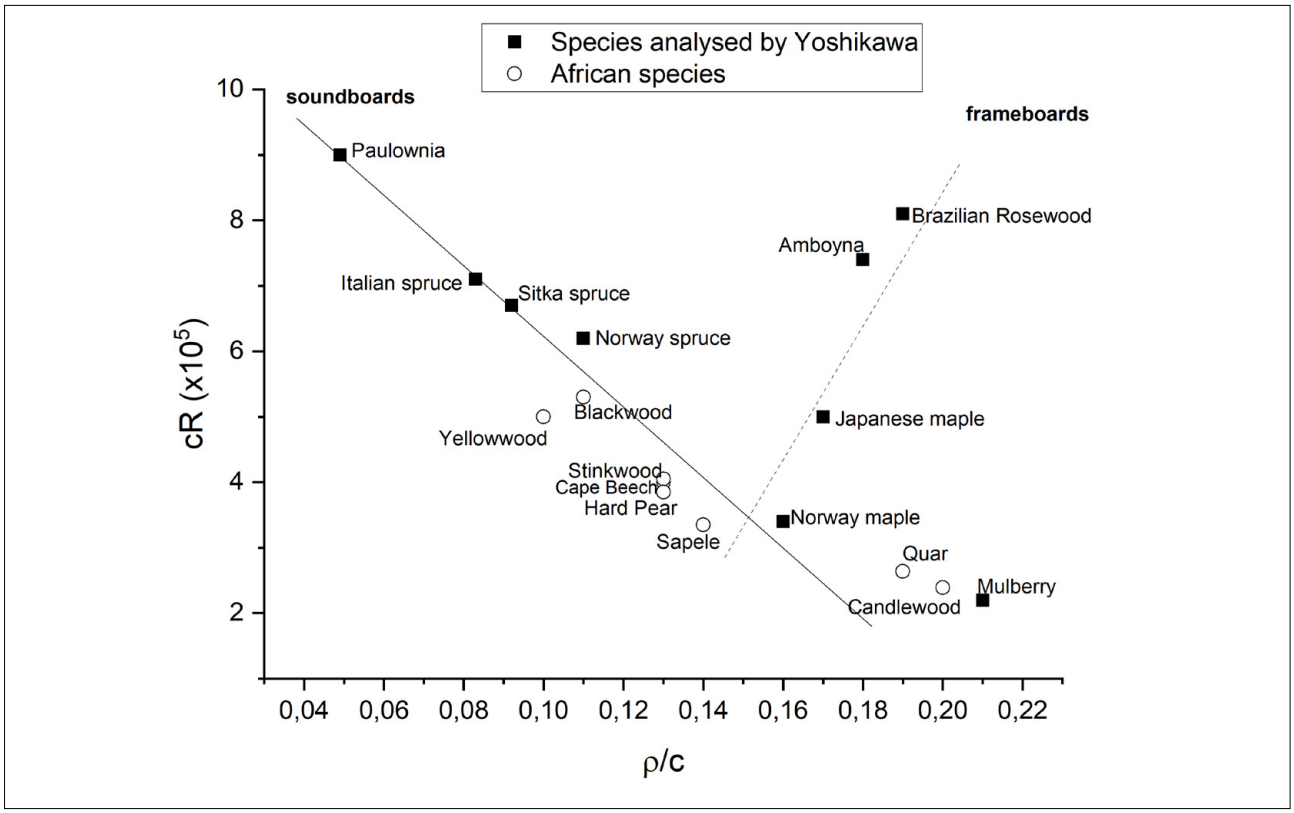



**Figure 1:** Acoustical classification suggested by Yoshikawa[4]. The regression lines indicate the properties of typical soundboards and frameboards, when the transmission parameter (cR) is plotted as a function of the antivibration parameter (ρ/c). Full squares indicate species reported in his study, while open circles represent the southern African species. Reference values sourced from Yoshikawa[4].

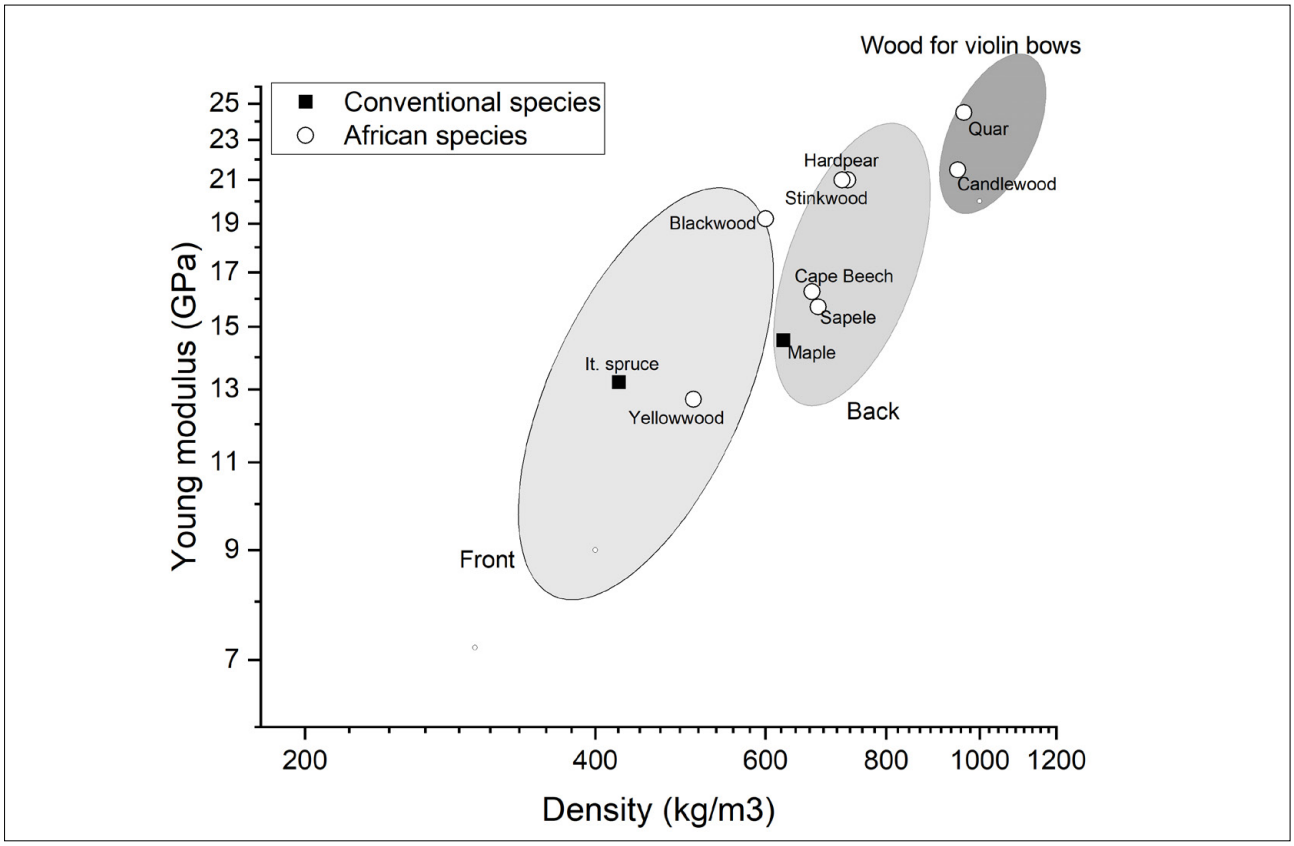



**Figure 2:** Classification suggested by Wegst[5]. The Young modulus E is plotted against the density ρ on a logarithmic scale. The highlighted areas indicate regions with suitable properties for soundboards (violin front), frameboards (violin back) and violin bows. Reference values sourced from Wegst[5].

**Figure 3:** The African violin made from yellowwood (soundboard) and sapele (remainder of the instrument).

A second violin will be made from blackwood for the soundboard, hardpear for the frameboard and sapele for the ribs and neck in a future project.

Due to the higher density of yellowwood and sapele compared to spruce and maple, the violin is a little heavier than conventional violins, weighing 540 g. In comparison, the conventional violin used for a comparison in this study weighs 458 g. All players noticed the extra weight, but did not experience it as a hindrance in playing the instrument.

### *Sound analysis*

The violin was played by various musicians on separate occasions, and all agreed that its sound is quite different from that of most other violins. The general assessment was that it has a very full, powerful sound with a strong lower register that projects well through the room. Given the lower radiation ratio of yellowwood compared to spruce, it was expected that it would not resonate as well as a conventional violin at higher frequencies.

Generally, the sound quality of an instrument can be described by the number of harmonics, or overtones. The more harmonics there are, the fuller the sound. Meinel[10] defined four frequency regions that add to the quality of sound:

1. The harmonics below 1 kHz should have high amplitudes, which results in a 'sonorous sound that carries well'.
2. The amplitudes around 1.5 kHz should be low, to prevent a 'nasal' character.
3. High amplitudes between 2 kHz and 3 kHz result in 'agreeable, bright' sound.
4. Small amplitudes for frequencies above 3 kHz result in a 'soft, pure' sound.

For the sound analysis, empty strings and the high E6 (an octave above the empty E string) were played for as long and as uniformly as possible and each note was repeated 10 times. The average power spectrum was computed for each note and used to compare the sound of the African violin with that of a 'conventional' violin that was comparable in all other aspects, except the wood species used for its construction. The same person played both violins with the same bow.

The normalised power spectral magnitudes were plotted as a function of the frequency and the peaks clearly show the dominant harmonics. Spectral power is the squared amplitude, which is typically represented on a logarithmic scale, as illustrated in Figure 4.

The frequency content is represented as a power spectrum, which is a standard representation in digital signal processing.[14] Plotting the spectral power rather than the amplitude (the square root of the power magnitude) of the complex spectrum makes smaller harmonics less visible and only the predominant harmonics appear as peaks. In addition, if the spectral power is plotted on a linear rather than the usual logarithmic scale, the main harmonic peaks are even more prominent, which makes a direct comparison easier.

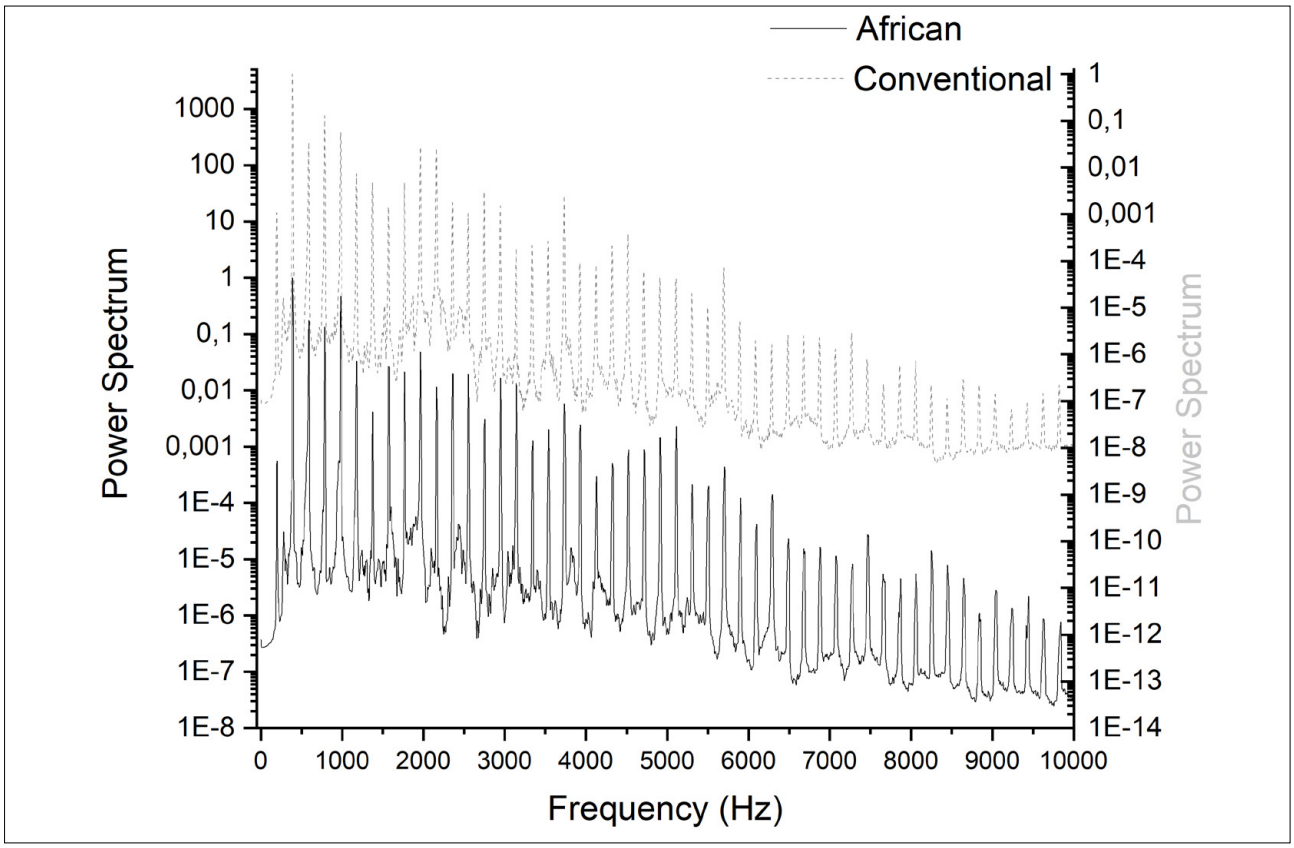



**Figure 4:** The sound signal (here G3) after Fourier transformation showing the normalised power spectrum on a logarithmic scale as a function of the frequency. The African violin is represented by the black line and the conventional violin by the dashed grey line.

The harmonic peaks continue throughout the high-frequency ranges and occur at multiples of the base frequency. In the linear representation, the frequency range was, however, limited to the values where significant peaks were present in the linear representation.

Figure 5a–e show the predominant harmonics in the linear spectra for both violins when the empty strings (G3, D4, A4, E5) and the octave above the empty E string (E6) were played. The fact that a linear scale is used means that only the most dominant peaks are clearly visible, while many smaller peaks disappear. In all graphs, the African violin is represented by a black line in the bottom half of the graph and the conventional violin by a dashed, grey line at the top half of the graph.

Both violins show higher amplitudes at low frequencies and smaller amplitudes around 1.5 kHz. However, the resonance frequencies of the African violin are visibly different from those of the conventional violin. It has stronger harmonics in the low-frequency range, which result in the full sound that carries well throughout the room. It also shows more overtones at higher frequencies. Especially for the G and D strings (Figure 5a and 5b), the frequency range between 2 kHz and 3 kHz, which is associated with a bright sound, is more pronounced for the African violin. However, harmonics in this frequency range are more pronounced for the conventional violin for the A and E strings (Figure 5c and 5d). For frequencies above 3 kHz, the African violin shows clearly higher amplitudes, which gives it a somewhat harsh sound, whereas the lower amplitudes of the conventional violin result in a soft sound.

The physical properties listed in Table 1 – especially the radiation ratio R, which is significantly lower in yellowwood than in spruce – suggest that the African violin would radiate sound less well in the high-frequency range. Contrary to this expectation, the African violin shows larger resonance peaks at high frequencies than does the conventional violin. This becomes even more evident if the amplitude rather than the power spectrum is plotted over a larger frequency range, as displayed in Figure 6, which shows the harmonics of the E string.

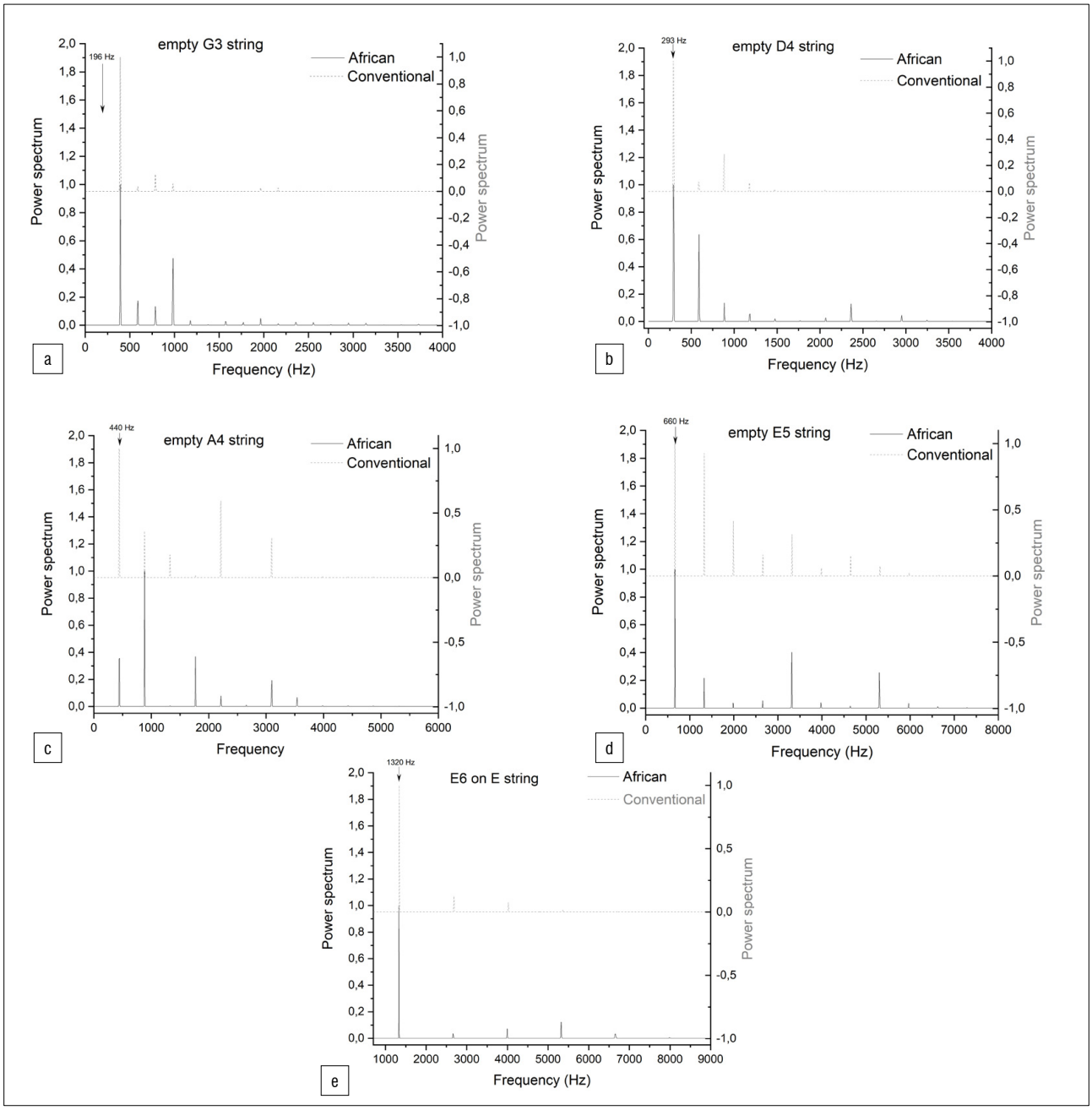



**Figure 5:** The predominant harmonics of both violins audible when the empty (a) G, (b) D, (c) A and (d) E strings are played, as well as the higher E6 (e). The base frequencies are indicated by an arrow.

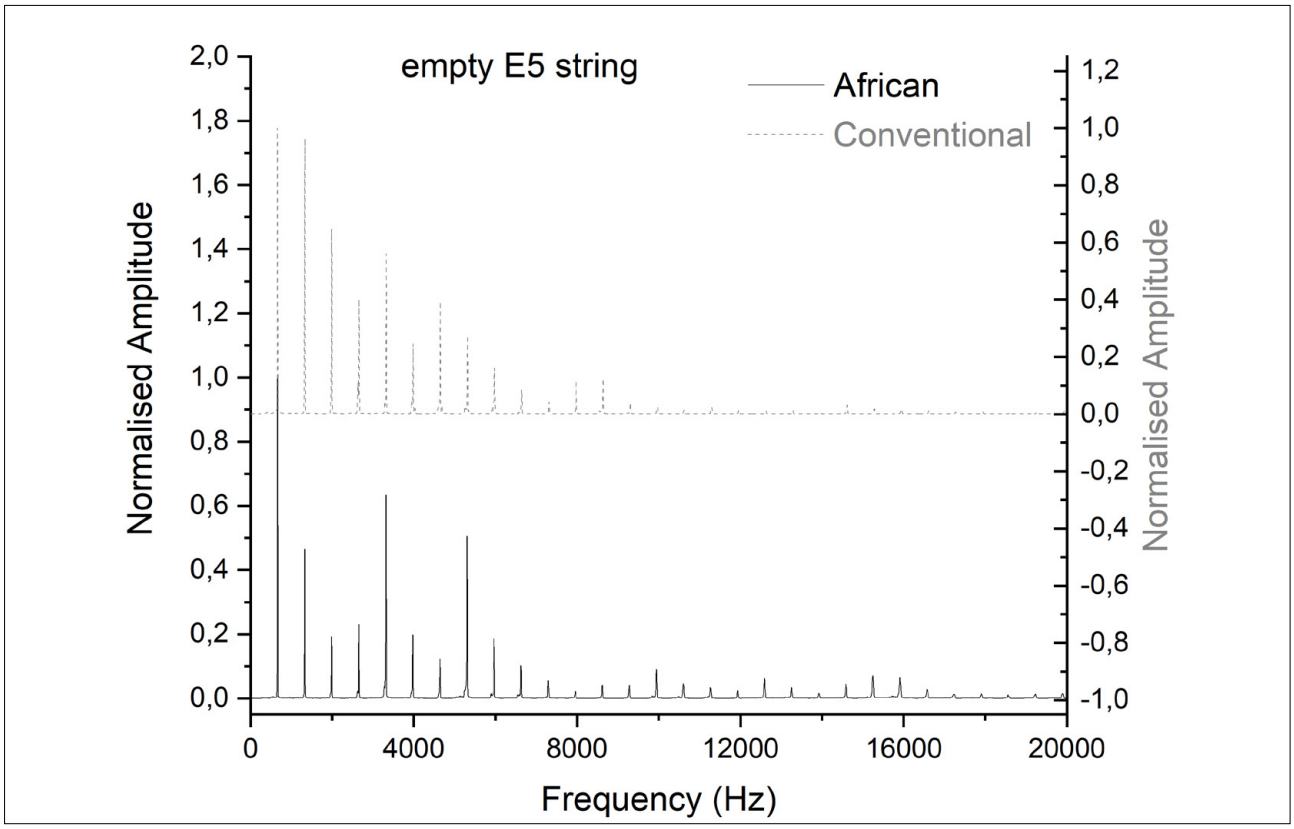



**Figure 6:** Normalised amplitude for both violins when the empty E string is played.

It can be seen that the amplitudes of the different harmonics below 8 kHz are different for the two violins, which results in the different sound. Surprisingly, the African violin shows significantly higher harmonics at frequencies above 10 kHz. This was not really expected, because yellowwood has a significantly lower radiation ratio (9.9) than the Italian spruce (13.2).

The aim of this project was to identify indigenous southern African wood species that are suitable as tonewood, and four species were identified: yellowwood and blackwood as possible soundboards and sapele and hardpear as possible frameboards. A first violin was made by a professional luthier from yellowwood and sapele and a second violin will be made as an ongoing project from blackwood and hardpear. Preliminary sound analysis showed that the African violin has a distinctly different sound with a powerful lower register and surprisingly many harmonics at higher frequencies. A more thorough sound analysis of both African violins will be performed in a future study.

## Conclusions

The acoustic properties of indigenous wood species from southern Africa were characterised and fitted into two established classification schemes used to identify suitable tonewoods for string instruments. Several wood species showed promise for use as soundboards and frameboards. Based on these results, a violin was made using yellowwood and sapele.

A comparative spectral analysis with a control instrument showed clear differences in the acoustic radiation pattern between the two violins. The African violin displayed larger harmonics at low and high frequencies for the G and D strings which result in a full sound that carries well through the room. For the higher notes – especially the E string and the E6 – high-frequency harmonics were lower in amplitude for the conventional violin, which results in a softer sound. Despite the differences, both violins were perceived to produce good sound quality. The difference in sound can be attributed to the wood species that were used to make the instruments and the results show that yellowwood and sapele are clearly suitable to be used as tonewoods and produce an instrument with a beautiful, though slightly different sound. The African violin has a significantly stronger sound in the lower frequencies that carries very well through the room. It also showed more harmonics with higher amplitudes in the high frequencies, which gives it a harsher sound than the conventional violin. Yellowwood and sapele can be used to make string instruments with a sonorous, strong sound.

## Acknowledgements

We thank Hannes Jacobs for producing a wonderfully unique violin and Gretna Heynike for providing the control violin and playing the two violins for the recording.

## Competing interests

We have no competing interests to declare.

## Authors' contributions

M.M.: Project management, conceptualisation, methodology, sample analysis, data collection, writing, student supervision. G.R.: Data collection, validation, writing. T.N.: Data processing and analysis, validation, writing.

## References

1. Richter HG. Holz als Rohstoff fuer den Musikinstrumentenbau [Wood as a raw material for making musical instruments]. Celle: Moeck; 1988. German.
2. Bucur V. Wood species for musical instruments. In: Acoustics of wood. 2nd ed. Heidelberg: Springer; 2006. p. 173–215. https://doi.org/10.1007/3-540-30594-7
3. Hutchins CM. The acoustics of violin plates. Sci Am. 1981;245:170–186. https://doi.org/10.1038/SCIENTIFICAMERICAN1081-170

4. Yoshikawa S. Acoustical classification of wood for string instruments. J Acoust. 2007;122:568–573. https://doi.org/10.1121/1.2743162
5. Wegst U. Wood for sound. Am J Bot. 2006;93:1439–1448. https://doi.org/10.3732/ajb.93.10.1439
6. Schelleng JC. Wood for violins. J Catgut Acoust Soc. 1982;37:8–19.
7. Barlow CY. Materials selection for musical instruments. Proc Inst Acoust. 1997;19:69–78.
8. Fritz C, Blackwell AF, Cross I, Woodhouse J, Moore BCJ. Exploring violin sound quality: Investigating English timbre descriptors and correlating resynthesized acoustical modifications with perceptual properties. J Acoust. 2012;131:783–794. https://doi.org/10.1121/1.3651790
9. Bissinger G. Structural acoustics of good and bad violins. J Acoust. 2008;124:1764–1773. https://doi.org/10.1121/1.2956478
10. Meinel H. Regarding sound quality of violins and a scientific basis for violin construction. J Acoust. 1957;29:817–822. https://doi.org/10.1121/1.1909064
11. Gabrielsson A, Jansson EV. Long-time-average-spectra and rated qualities of twenty-two violins. Acustica. 1979;42(1):47–55.
12. Jansson EV. Admittance measurements of 25 high quality violins. Acta Acoust. 1997;83:337–341.
13. Dünnwald H. Deduction of objective quality parameters on old and new violins. J Catgut Acoust Soc. 1991;1(7):1–5.
14. Rabiner L, Gold B. Theory and application of digital signal processing. Englewood Cliffs, NJ: Prentice Hall Inc; 1975.