



South African *Journal of Science*

volume 114
number 9/10



Food security is key to
achieving UNAIDS targets

Vancomycin-resistant
enterococci in South Africa

First ground-based
recording of elusive sprites
over southern Africa

Publication incentives
based on journal rankings
disadvantage local journals





eISSN: 1996-7489

EDITOR-IN-CHIEF

John Butler-Adam 
Academy of Science of South Africa

MANAGING EDITOR

Linda Fick 
Academy of Science of South Africa

**ONLINE PUBLISHING
SYSTEMS ADMINISTRATOR**

Nadine Wubbeling 
Academy of Science of South Africa

**ONLINE PUBLISHING
ADMINISTRATOR**

Sbonga Dlamini 
Academy of Science of South Africa

ASSOCIATE EDITORS

Nicolas Beukes
Department of Geology, University
of Johannesburg

Chris Chimimba
Department of Zoology and
Entomology, University of Pretoria

Linda Chisholm
Centre for Education Rights and
Transformation, University of
Johannesburg

Teresa Coutinho
Department of Microbiology and
Plant Pathology, University of
Pretoria

Tania Douglas 
Division of Biomedical Engineering,
University of Cape Town

Menán du Plessis 
Department of General Linguistics,
Stellenbosch University

Maryna Steyn 
School of Anatomical Sciences,
University of the Witwatersrand

Pieter Steyn
Department of Chemistry
and Polymer Science,
Stellenbosch University

Marco Weinberg
Department of Molecular Medicine
and Haematology, University of
the Witwatersrand

Patricia Whitelock
South African Astronomical
Observatory

**ASSOCIATE EDITOR
MENTEES**

Maitumeleng Nthontho
Department of Education
Management and Policy Studies,
University of Pretoria

Yali Woyessa
Department of Civil Engineering,
Central University of Technology

Leader

And out into the world they go ... research on the international stage

John Butler-Adam 1

Obituary

Claire Penn: Speech-language pathologist, teacher, supervisor, researcher and
mentor (1951–2018)

Jennifer Watermeyer 2

Book Review

The Square Kilometre Array radio telescope: A media darling

Ian S. Glass 3

What kind of people have we formed in our university graduates?

Chrissie Boughey 4

Commentary

Using the curriculum to enhance teaching and learning

Chrissie Boughey 5

Searching for David within the Goliath of alien woody plant invasions in the
Western Cape Province

Anthony J. Mills & Jessica L. Allen 8

The ecology of drought – a workshop report

*Anthony M. Swemmer, William J. Bond, Jason Donaldson, Gareth P. Hempson,
Johan Malherbe & Izak P.J. Smit* 11

Publication incentives based on journal rankings disadvantage local publications

Alan T.K. Lee & Carol A. Simon 14

Does South African research output promote innovation?

Roula Inglesi-Lotz & Anastassios Pouris 17

Review Article

Emergence of vancomycin-resistant enterococci in South Africa: Implications
for public health

Frank E. Tatsing Foka, Ajay Kumar & Collins N. Ateba 20

Aetiological doctrines and prevalence of pellagra: 18th century to middle
20th century

Margaretha Viljoen, Priyesh Bipath & Johannes L. Roos 27

EDITORIAL ADVISORY BOARD

Laura Czerniewicz 
Centre for Higher Education
Development, University of Cape Town

Roseanne Diab 
Academy of Science of South Africa

Hassina Mouri
Department of Geology,
University of Johannesburg

Johann Mouton
Centre for Research on Science and
Technology, Stellenbosch University

Sershen Naidoo
School of Life Sciences, University of
KwaZulu-Natal

Maano Ramutsindela
Department of Environmental &
Geographical Science, University of
Cape Town

Published by

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

Design and layout

SUN MeDIA Bloemfontein
T: 051 444 2552
E: admin@sunbloem.co.za

Correspondence and enquiries

sajs@assaf.org.za

Copyright

All articles are published under a
Creative Commons Attribution Licence.
Copyright is retained by the authors.

Disclaimer

The publisher and editors accept no
responsibility for statements made by
the authors.

Submissions

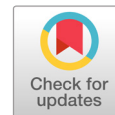
Submissions should be made at
www.sajs.co.za

Research Article

Monthly average daily solar radiation simulation in northern KwaZulu-Natal: A physical approach <i>Betty Kibirige</i>	34
Food insecurity, HIV status and prior testing at South African primary healthcare clinics <i>Makandwe Nyirenda, Renee Street, Tarylee Reddy, Susie Hoffman, Suraya Dawad, Kelly Blanchard, Theresa M. Exner, Elizabeth A. Kelvin, Joanne E. Mantell & Gita Ramjee</i>	42
Alcohol and other drug use among women seeking substance abuse treatment in the Western Cape, South Africa <i>Siphokazi Dada, Nadine Harker Burnhams, Ria Laubscher, Charles Parry & Bronwyn Myers</i>	48
First report of bacterial endophytes from the leaves of <i>Pellaea calomelanos</i> in South Africa <i>Siphiwe G. Mahlangu & Mahloro H. Serepa-Dlamini</i>	55
Mapping chlorophyll-a concentrations in a cyanobacteria- and algae-impacted Vaal Dam using Landsat 8 OLI data <i>Upa E. Malahlela, Thando Oliphant, Lesiba T. Tsoeleng & Paidamwoyo Mhangara</i>	64
Application of science and technology by the South African food and beverage industry <i>Lisa-Claire Ronquest-Ross, Nick Vink & Gunnar O. Sigge</i>	73
First ground-based observations of sprites over southern Africa <i>Stanislaus Nnadih, Mike Kosch, Peter Martinez & Jozsef Bor</i>	84

Cover caption

In an image captured from
the International Space Station, a
red sprite can be seen above a lightning flash
(image: NASA). In an article on page 84, Nnadih and
colleagues present the first ground-based recordings of sprites over
southern Africa.



And out into the world they go ... research on the international stage

The Leader in the July/August issue of the *South African Journal of Science* reflected briefly on the loss of senior university managers who have left the local system to take up leading positions in universities elsewhere in the world. The Leader mentioned that this has also been true of scientists and scholars – and in this Leader we reflect on a different aspect of internationalisation – changes in patterns of research output and cooperation.

To do so, the Leader relies substantially on the June 2018 issue of *SciBytes*¹, published by the DST-NRF Centre of Excellence in Scientometrics and Science, Technology and Innovation Policy (SciSTIP). The authors of the June issue, Prof. Johann Mouton and Dr Jaco Blanckenberg, have granted permission to draw on their work here.

In making their calculations, the authors make use of the Web of Science database, and confine their assessment to South Africa's publications in two categories: 'articles' and 'review articles'. This means that they exclude documents such as books, book chapters and conference proceedings in their counts which are based on three indicators: (1) publication output, (2) international collaboration and (3) citation visibility or impact.

In the 17-year period between 2000 and 2016, three major shifts have taken place in the nature of South African scientific publishing. First, research publications have not just increased (as might be expected) but increased significantly. Secondly, South African scientists have come to collaborate considerably more frequently with scientists and scholars internationally than before. Thirdly, the growth in South Africa's publication output has taken place at the same time as there has been an increase in the visibility of our scientific papers.

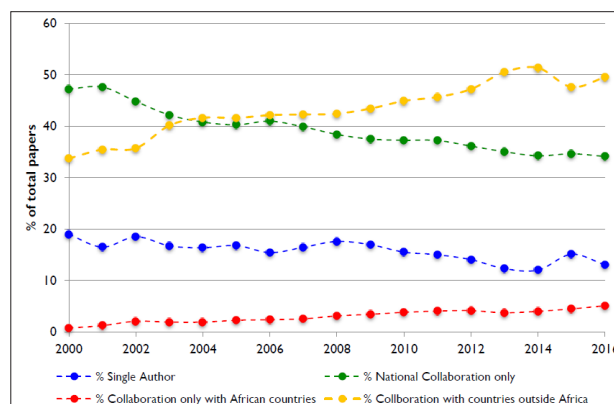
South Africa's publication output in the Web of Science has increased from 3668 publications in 2000 to 15 550 in 2016. This increase translates into an average annual growth rate of 2.9%. This growth has been such that South Africa's share of world output more than doubled from 0.4% in 2000 to 0.91% in 2016. Not surprisingly, these results have translated into an improved position when comparing South Africa with other countries. As far as country rank is concerned, South Africa has improved its ranking in the world from position number 34 in 2000 to 28 in 2016, despite obvious competition.

It is standard bibliometric practice to measure research collaboration by looking at patterns of co-authorship in scientific papers. The *SciBytes* report followed the same practice and specifically distinguished between four categories of collaboration:

- No collaboration (either single-authored articles or single-institution authorship)
- National collaboration (multiple authors from more than one institution in South Africa)
- International collaboration with scientists from African countries only
- International collaboration with scientists from countries outside of Africa

The results (Figure 1) show a clear trend towards more international collaboration. This in itself is a desirable development as increased

international collaboration often translates into higher citation impact, increases in networks and access to more funding opportunities.



Source: Mouton and Blanckenberg¹

Figure 1: South Africa's publication collaboration profile (2000–2016).

In 2000, about a third of South Africa's papers involved co-authorship with at least one foreign author. By 2016, this proportion had increased to 50%. The increase in international collaboration has occurred at the 'expense' of national collaboration (which declined from 47% to 34% over the same period) as well as a clear decline in single-authored publications. There is a small, but steady, trend of increasing collaboration with scientists and scholars in the rest of Africa: this proportion increased from a near-zero base in 2000 to 5% in 2016.

As for the third change, an increase in scientific output does not necessarily imply that such output is recognised by other scientists working in the same fields. So the question was asked as to whether or not the substantial growth in South Africa's production of scientific papers translated to an increased visibility amongst scientists.

The visibility of science is partially captured by looking at the number of times research publications are cited in the publications of other researchers. Citation practices were normalised in order to allow for comparison. The resulting analyses show that the citation impact of South Africa's scientific papers has increased steadily from 0.8 in 2000 to 1.1 in 2016. This is a very positive result as a score of above 1 means that South Africa's papers are, on average, being cited slightly more than all the papers in the fields in which we publish.

In short – the 17-year period saw a noteworthy increase in research publications; there has been a significant increase in internationally cooperative research and publishing; and papers involving South African scientists are being cited more frequently (as might be expected as a result of international cooperation) than previously. There is much to be pleased about and to support going forward.

Reference

1. Mouton J, Blanckenberg J. How well is South African science doing? *SciByte@SciSTIP*. 2018 June No 2 [cited 2018 Aug 14]. Available from: http://www0.sun.ac.za/scistip/wp-content/uploads/2018/06/SciBytes@SciSTIP_02.pdf





Claire Penn: Speech-language pathologist, teacher, supervisor, researcher and mentor (1951–2018)

AUTHOR:

Jennifer Watermeyer¹

AFFILIATION:

¹Health Communication Research Unit, Department of Speech Pathology and Audiology, School of Human and Community Development, University of the Witwatersrand, Johannesburg, South Africa

CORRESPONDENCE TO:

Jennifer Watermeyer

EMAIL:

Jennifer.Watermeyer@wits.ac.za

HOW TO CITE:

Watermeyer J. Claire Penn: Speech-language pathologist, teacher, supervisor, researcher and mentor (1951–2018). *S Afr J Sci.* 2018;114(9/10), Art. #a0283, 1 page. <https://doi.org/10.17159/sajs.2018/a0283>

PUBLISHED:

11 September 2018

Professor Emeritus Claire Penn passed away peacefully on 21 July 2018 after a long illness. Claire dedicated a significant portion of her life – 44 years – to the University of the Witwatersrand. She received her degree in logopaedics (speech and hearing therapy) cum laude in 1972.

Claire held the endowed chair of Speech Pathology and Audiology and was Director of the Health Communication Research Unit in the School of Human and Community Development at Wits University. She was the first A-rated scientist in the Faculty of Humanities. She had a strong international profile, served on the editorial board of a number of international journals, was a member of several international speech pathology organisations, and regularly delivered international invited keynote addresses. Claire published over 100 papers in local and international peer-reviewed scientific journals, 25 chapters and 4 books (including an extensive multi-volume dictionary on southern African signs).

Claire was instrumental in the development of the field of health communication in South Africa. The merging of her lines of research afforded a rich and mutually beneficial interface with societal impact and recognition. Her research strongly suggested the need for an expanded role for the profession of speech-language therapy, both in national and international contexts. Her work also consistently demonstrated the major contribution that South African research findings and methods and insights have to offer the global forum.

Claire was a recipient of a number of awards, including the Order of Mapungubwe (Silver) in 2007, for her 'excellent contribution to the field of speech and language pathology, especially in the area of linguistics, sign language, child language, aphasia and head injury and producing ground-breaking research in understanding the complexities of human communication'. She was the winner of the Science & Technology category of the Shoprite/Checkers SABC2 Woman of the Year award in 2008 and the Department of Science and Technology Distinguished Woman Scientist award (academic excellence in Social Sciences or Humanities) in 2010. Claire was a member of the Academy of Science of South Africa.

Claire had a brilliant and curious mind; she had tremendous vision and a gift for seeing the big picture. She was ahead of her time in many ways – teaching contextually relevant ways of doing therapy and research right from the start of her career. Claire set high academic standards for herself and others. She built strong academic ties both locally and internationally. She broke through glass ceilings and fought for the place of women in academia. She was a fierce defender of ethical principles and patients' rights, especially those patients marginalised by communication disorders and language barriers. She taught her students to be advocates and activists. Claire was never more at home than when sitting in a clinic somewhere in a rural setting, doing ethnography under a tree and interviewing patients, genuinely interested in improving their quality of life and care.

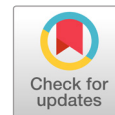
Claire will be remembered as an outstanding teacher, supervisor, researcher and mentor. She had great vision for the profession and she has inspired generations of speech therapists and audiologists. Many of those who were taught and mentored by Claire have gone on to become research and clinical leaders, both locally and internationally.

Claire is survived by her partner, Martin Templer, and son, Adam Penn-Nicholson.



Photo: Joanne Neille

© 2018. The Author(s).
Published under a Creative
Commons Attribution Licence.

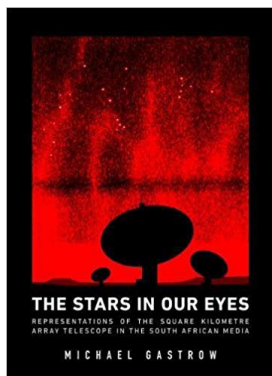


The Square Kilometre Array radio telescope: A media darling

BOOK TITLE:

The stars in our eyes:
Representations of the Square
Kilometre Array telescope in the
South African media

BOOK COVER:



AUTHOR:

Michael Gastrow

ISBN:

9780796925471 (paperback);
9780796925442 (ebook)

PUBLISHER:

HSRC Press, Cape Town;
ZAR250

PUBLISHED:

2017

REVIEWER:

Ian S. Glass

AFFILIATION:

South African Astronomical
Observatory, Cape Town,
South Africa

EMAIL:

isg@sao.ac.za

HOW TO CITE:

Glass IS. The Square Kilometre
Array radio telescope:
A media darling. S Afr J
Sci. 2018;114(9/10), Art.
#a0284, 1 page. [https://doi.
org/10.17159/sajs.2018/a0284](https://doi.org/10.17159/sajs.2018/a0284)

PUBLISHED:

11 September 2018

Michael Gastrow is a member of the Education and Skills Development Programme of the Human Sciences Research Council (HSRC), with a special interest in public relations with science. The subject of his book is the manner in which the Square Kilometre Array (SKA) radio telescope has been presented to the South African public. Enjoying strong political support and as the most expensive scientific project to be constructed in the country, the SKA has been something of a media darling.

Although the international SKA project began in 2000 and construction has yet to start, the book confines itself to a very brief interval in its history, namely September 2011 to August 2012, during which time the decision was reached to site the instrument partly in the Karoo and partly in Western Australia.

The data for Gastrow's thesis-like book comprise 174 articles from print and online media together with 1588 tweets and 27 interviews. These form a rather small core of information that is examined from every conceivable angle. The interviews mentioned were with spokespersons for the government agencies involved, a mere seven science journalists, various managers, public relations people within the SKA project and relevant university staff.

The placing of a scientific story in the print media by a science journalist is the result of a two-step process. Firstly, information has to be accumulated, usually by interaction with individual scientists and organisational spokespersons. The written article then must get over a serious hurdle which is the 'gatekeeping' that takes place in the newsroom: an editor must decide if a story is going to have enough impact to be worth publishing. Of course, tweeting in the electronic media is a less formal route but one that usually has less public impact.

Although the South African media have presented the SKA as a largely South African project, it is very much an international project, with its headquarters at Jodrell Bank in England. The competition, effectively between South Africa and Australia, over its location was easily the most interesting source of news stories so far as local editors were concerned – the leaked details of the penultimate decisions on this issue and the last minute politicking on how the split was to be made between the two countries provided plenty of grist for the editorial mills.

Various peripheral issues have, however, received relatively little attention in the press. The *Astronomy Geographic Advantage Act (2007)*, strongly supported by the South African government, was a potential 'hot potato', impinging as it does on fracking interests and effectively banning cellphones and petrol-engined vehicles from the neighbourhood of the telescope. Local farmers and their workers have been vocal about their concerns and these concerns have had to be taken seriously.

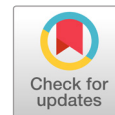
Why this particular project resonated so successfully with the ANC government is a question that has frequently been posed and is also considered here. Gastrow quotes the political scientist Keith Gottschalk who suggested at a symposium of the Astronomical Society of Southern Africa in 2005 that national prestige, the dignity of the African continent and black dignity itself were the considerations with the greatest political appeal. Other justifications frequently mentioned are the technological benefit that the development of the equipment will bring to the country and its aspect as a beacon to attract interest in science and engineering in future generations. This point of view echoes somewhat a widely quoted 1996 White Paper of the then Department of Arts, Culture, Science and Technology:

Scientific endeavour is not purely utilitarian in its objectives and has important associated cultural and social values. It is also important to maintain a basic competence in 'flagship' sciences such as physics and astronomy for cultural reasons. Not to offer them would be to take a negative view of our future – the view that we are a second class nation, chained forever to the treadmill of feeding and clothing ourselves.

The book is quite unusual from a scientist's point of view, in that it is not concerned with the SKA per se, but rather with its treatment in the media. Strictly speaking, the construction of the SKA is not science, but an engineering exercise that is expected to yield a telescope with an unprecedented capability that may reasonably be expected to lead to new insights and discoveries in the fullness of time. Such outcomes have normally been the experience as telescopes of larger and larger diameters have been constructed, as also with more and more powerful particle accelerators.



© 2018. The Author(s).
Published under a Creative
Commons Attribution Licence.



What kind of people have we formed in our university graduates?

BOOK TITLE:

Going to university: The influence of higher education on the lives of young South Africans

BOOK COVER:



AUTHORS:

Jennifer M. Case, Delia Marshall, Sioux McKenna and Disaapele Mogashana

ISBN:

9781928331698 (softcover);
9781928331704 (open access
ebook)

PUBLISHER:

African Minds, Cape Town;
ZAR200 (softcover)

PUBLISHED:

2018

REVIEWER:

Chrissie Boughey 

AFFILIATION:

Centre for Higher Education
Research, Teaching and Learning
(CHERTL), Rhodes University,
Grahamstown, South Africa

EMAIL:

C.Boughey@ru.ac.za

HOW TO CITE:

Boughey C. What kind of
people have we formed in our
university graduates? S Afr
J Sci. 2018;114(9/10), Art.
#a0287, 1 page. [https://doi.
org/10.17159/sajs.2018/a0287](https://doi.org/10.17159/sajs.2018/a0287)

PUBLISHED:

11 September 2018

© 2018. The Author(s).

Published under a Creative
Commons Attribution Licence.

Globalisation has had an enormous impact on all who work or study in higher education today. Discourses constructing the role of universities as providers of the 'knowledge workers' needed to fuel the global economy have impacted on the expectations of students, on the curricula with which they engage, on the work of academic staff as well as on the 'management teams' that now make key decisions. Other discourses, associated with what is widely termed 'New Public Management', focus our attention on efficiency and accountability. Of course, state-funded universities need to be accountable and efficient, especially in a country like South Africa where the demands on the public purse are many and, often, desperate. Problems arise, however, when efficiency and accountability discourses lose sight of the fact that higher education involves human beings. When this happens, we focus on the 'production of outputs', where the term 'output' refers not only to artefacts such as publications in the form of journal articles and books but to the young people passing through our universities.

In all this, *Going to University*, authored by Jenni Case, Delia Marshall, Sioux McKenna and Disa Mogashana, offers a very different understanding of what our universities can do and of the experiences of their students. The book does this through drawing on 73 in-depth interviews with students. All of the student participants in the study had entered one of three universities – the University of Cape Town, the University of the Western Cape or Rhodes University – in 2009. By the time the research project on which the book is based began in 2015, 6 years had passed. The vast majority of interviewed students had succeeded in attaining degrees. These were not always the degrees they had imagined when they enrolled and not always at the university at which they began their academic careers. Some had 'dropped out' – victims of the 'attrition' that causes so much alarm and which occupies many individuals, particularly 'managers of teaching and learning, with finding solutions. However, what the book does, and what the 'big data' of cohort analyses never could do, is illuminate the richness of the experience of these individuals in a way which ultimately must give us hope and which offers a very different perspective on the gloomy picture most often presented of our higher education system.

The problem of 'attrition', or 'dropping out', is often associated with concepts such as 'wastage'. The book reveals very clearly that the lives of the young people participating in the study who dropped out of university certainly were not 'wasted'. The interviews show that the experience of being in a university gave these young people confidence, allowed them to draw on networks they otherwise would not have accessed and, in cases in which they wanted to, the means to eventually complete a degree at another institution.

Although the book is very accessible to a wide range of readers, it is not 'light' on theory. The authors draw on complex theory, more specifically the work of Archer¹, in the design of their small-scale study based on in-depth interviews. While the use of this theory is rigorous, it is never intrusive to the reader, not least because details of the methodological approach are included as an appendix rather than as a chapter that 'interrupts' the stories. Some of the theory used to support analyses and explanations of the data – most notably in the area known as the 'sociology of knowledge' including, for example, the work of Bernstein² and Maton³ – is notoriously difficult to access. However, the use of interview data to illuminate concepts (for example, that the way knowledge is structured in different disciplinary areas can make progress in the humanities different from that in the sciences), means that the book functions as an extremely useful resource for the increasing number of postgraduate students using this type of theoretical approach.

Other merits notwithstanding, what makes the book especially significant at this time in our history is the hope it offers. This review began by pointing to the pessimism surrounding higher education in South Africa, a phenomenon which was arguably exacerbated by the student protests of 2015 and 2016. Although the protests elicited very negative reactions from many sectors, the protesters' descriptions of their experiences in our universities as 'suffering' has to make us think about what is going wrong and what is going right. *Going to University* shows us that we are getting some things right as well as what we could do to make things better.

In the context of all the negativity around higher education, perhaps most important is the authors' response to a question they ask at the end of the book: 'What kind of people have we formed in our university graduates?'

Their response is that, according to their study, the young people who have been through our universities '...are independently minded and socially progressive. They are getting traction in their careers and they are acting with thoughtfulness and responsibility. They are thinkers and they mostly engage critically with the world and their place in society. Many are aware of inequalities in society and of their own experiences of privilege.'

Going to University is important reading for anyone who cares about higher education in South Africa. The fact that the book is published by African Minds and, thus, is open access is but one more reason for it to be read widely.

References


1. Archer M. Making our way through the world: Human reflexivity and social mobility. Cambridge: Cambridge University Press; 2000. <https://doi.org/10.1017/CBO9780511618932>
2. Bernstein B. Pedagogy, symbolic control and identity. 2nd ed. New York: Rowman & Littlefield; 1996.
3. Maton K. Knowledge and knowers: Towards a realist sociology of education. Abingdon: Routledge; 2013. [https://doi.
org/10.4324/9780203885734](https://doi.org/10.4324/9780203885734)





Using the curriculum to enhance teaching and learning

AUTHOR:

Chrissie Boughey¹ 

AFFILIATION:

¹Centre for Higher Education Research, Teaching and Learning (CHERTL), Rhodes University, Grahamstown, South Africa

CORRESPONDENCE TO:

Chrissie Boughey

EMAIL:

C.Boughey@ru.ac.za

KEYWORDS:

higher education; curriculum reform; attrition; outcomes-based approaches

HOW TO CITE:

Boughey C. Using the curriculum to enhance teaching and learning. *S Afr J Sci.* 2018;114(9/10), Art. #a0288, 3 pages. <https://doi.org/10.17159/sajs.2018/a0288>

PUBLISHED:

11 September 2018

The publication of cohort studies tracking students' performance has given rise to concerns about the efficiency of the South African higher education system at a number of levels. These studies, which began with Scott et al.'s¹ pioneering work in 2007 and which have continued with the CHE's² annual *Vital Stats* series, show that, regardless of the university at which they are enrolled, the subject area or the type of qualification for which they are registered, black South Africans fare less well than their white peers. At institutional levels, alarm at such observations is seen in efforts to manage success, throughput and drop-out rates through the appointment of key individuals such as deans and deputy vice chancellors responsible for teaching and learning. At a national level, concern about the performance of the system overall has led to the introduction of, first, Teaching Development Grants and, now, University Capacity Development Grants. Both of these mechanisms provide earmarked funding aimed at enhancing the quality of teaching and learning.

Although practitioners working in what is often termed the 'Academic Development Movement' have long identified the centrality of the curriculum in efforts to improve teaching and learning,³ problematic 'common sense' tends to prevail with the result that curriculum reform is neglected at the expense of ad hoc initiatives which take place outside mainstream learning.

This neglect is in spite of the fact that South Africa has been in a state of almost continuous curriculum reform for more than two decades. Curriculum reform in South Africa began with the registration of qualifications on the National Qualifications Framework (NQF) following the establishment of the South Africa Qualifications Authority in 1995.⁴ At the same time as the requirement to register qualifications on the NQF arose, the country was engaging with the need to participate in the global economy after years of apartheid-related isolation. As Kraak⁵ points out, in the early 1990s, the South African labour movement was acutely aware of the need to engage with the global economy if even higher levels of unemployment were to be avoided. As a result, the unions were happy to endorse the development of vocationally oriented programmes in higher education in order to produce the 'knowledge workers' constructed as critical to participation in the global economy by dominant discourses.

Although the former 'technikons' had always offered vocational programmes, the thinking at the time when the need to register qualifications arose led to many traditional universities moving towards offerings with a stronger workplace focus. In some cases, this resulted in wide-scale reorganisation of academic departments into schools and other structures. The emergence of three distinct types of institution as a result of the processes leading to the National Plan for Higher Education⁶ in 2002 also meant that many more institutions moved towards the development of vocational programmes, particularly given the series of mergers and incorporations at that time which saw many 'traditional' universities merging with technikons to become 'comprehensive universities'.

The registration of qualifications on the NQF is dependent on the use of the 'learning outcome' as an organising principle. Learning outcomes describe what learners will be able to do upon completion of a programme of study and are typically expressed in terms of knowledge, skills and 'attitudes'. The construct of the learning outcome thus means that a common language necessary for the functioning of the framework is available to describe qualifications as they are registered.

Vocational programmes provide students with the knowledge necessary to address the learning outcomes around which they are designed. As a result of the focus on skills and the provision only of knowledge deemed necessary for a vocational area, mastery of a coherent body of disciplinary knowledge at undergraduate level has often become neglected in all but the few universities that continue to offer the general formative degree. These degrees allow students to engage with a number of different subject areas as they enter undergraduate study and to build coherent understandings of the 'knowledge structures' of the subjects they carry through to third year as 'majors'.

More than 20 years on from the time South African universities began to engage with the development of vocational programmes, the emergence of work produced in the field known as the 'sociology of knowledge', particularly that of Wheelahan⁷ and Maton⁸, challenges the thinking that has influenced curriculum reform in South Africa so heavily. These researchers argue that bodies of disciplinary knowledge are 'powerful' in that they allow knowers to move beyond the specific context for which they were trained, and for which they have mastered only the knowledge necessary to meet its demands, to other contexts. Critically, they argue that mastery of theoretical knowledge, often neglected in the focus on skills in outcomes-based approaches, allows knowers to imagine worlds that do not yet exist. Disciplinary knowledge, in the form of a cumulative 'knowledge structure', is thus seen as central to innovation and social change. Providing students with only the knowledge necessary to deal with problems in a specific context, as is the case in most vocational programmes, is seen to be particularly limiting given the pace at which the world of work is developing. Arguably, many of the jobs for which students have been trained in narrow vocational programmes will cease to exist before the end of their working lives.

The book reviewed elsewhere in this issue, Case et al.'s⁹ *Going to University*, shows that the vocationally oriented programmes introduced at so many universities have another disadvantage, in that they 'lock' students into a specific programme of study. Much as a career in a particular area might have seemed attractive when students were applying to study at university, the realities of studying are often very different. As one student cited in the book notes, 'I wanted to do astrophysics when I came here, I was intent on landing on the moon, but then you get here and realise that astrophysics is actually just physics'.

Other students in the study underpinning the book found that they were suddenly 'passionate' about a subject they encountered in the course of their studies, possibly an area they knew nothing about when they were at school. One student, for example, initially wanted to do Clinical Psychology but then, '...got introduced into Biological Psychology and I fell in love and it was like, "Oh my goodness, I don't want to do Clinical Psychology anymore". I'm like over it.'

The curriculum structure of a general formative degree that allows students to engage with a range of subject areas before following two of these through as 'majors' means that students can redirect themselves in ways that draw on their interests or 'passions' if their original choices fail to meet their expectations. It is not difficult to see how studying a subject that a young person feels is truly interesting would have a positive effect on academic performance.

Flexibility in the curriculum offers another critical advantage. In a tightly structured vocational programme, failing a course may mean that the student needs to repeat a year or even, if failures are repeated, to begin a completely new programme in order to attain a qualification. A flexible structure allows students to redirect themselves and find alternative routes through the curriculum without necessarily taking extra time.

In spite of the wealth of research demonstrating its centrality, the curriculum itself tends to be neglected as a means of enhancing teaching and learning. The most recent round of curriculum reforms following the publication of the Higher Education Qualifications Sub-Framework¹⁰ in 2013 have arguably led to little more than technical adjustments to curricula as institutions comply with the requirements of the framework rather than to considerations of the way curricula could impact more positively on students' experiences and, thus, to the performance of the system overall.

It is also enormously disappointing that alternatives to the 'vocational route' to the LLB qualification have been ignored in the recent CHE reaccreditation processes. Many students come into our universities thinking they want to become lawyers, or under pressure from families to become lawyers. They then often find that law is not for them, either because they fail or because the abstract study of the law in the early undergraduate years, that forms the basis for the more practice-based work that comes later, is something other than what they imagined. A route towards an LLB that consists of a 3-year general formative degree with the 2-year professional 'top up' of the law degree itself allows such students to find other pathways to a qualification and a career and, often, their passion along the way, if things go wrong. The emphasis on the 4-year LLB in the CHE accreditation processes thus has the potential to impact on the aspirations of many students to achieve a qualification if they do indeed find the law is not for them.

Curriculum reform aimed at enhancing teaching and learning could mean more than acknowledging the disadvantages of narrowly focused vocational routes. Universities have had 'Extended Programmes with an Integrated Phase' since the introduction of Foundation Programme Grants by the DHET¹¹ in 2004. These programmes allow for the insertion of up to 120 additional 'developmental' credits in a regular accredited programme, which is then 'extended' by up to a year in order to allow for the extra tuition and study. Developmental credits do not count towards the qualification associated with the programme into which they have been inserted. Rather, the intention is that they provide the support necessary for students to engage with the 'regular' credits it requires. The DHET policy¹¹ allocating funding for the credits allows for these developmental credits to be inserted at any point in the extended programme, although the most common practice is to locate them within the first 2 years.

Over time, a number of course types have been developed in order to accommodate these credits. These are:

- Fully foundational courses in which all the learning is aimed at filling the gaps in knowledge deemed necessary as a basis for tertiary study.
- Extended courses in which the time taken to complete a course is doubled to allow for the insertion of development credits. This

might mean that a semester long course is run over an entire year with extra tuition being offered within the course.

- Augmented courses in which additional material is inserted into the regular course alongside the regular tuition time needed to complete it. This means that a semester course, offering, say, four contact periods per week, would continue to run over the length of a semester but might offer six or eight contact periods to cover the developmental credits.

It is not difficult to see how courses in regular programmes with high failure rates could be adapted to draw on these course types. Often the courses with the highest failure rates are 'gateway' courses, so called because they must be passed if a student is to proceed further. In my university, for example, a first-year 15-credit course in Cell Biology is a 'gateway' to routes through the (already flexible, general formative) curriculum in the natural sciences. The failure rates in this course were higher than those of other courses. Using funding from the Teaching Development Grant and, now, the University Capacity Development Grant, the course was developed into an 'augmented course' and a specialist in the area with a particular interest in teaching and learning was employed to teach the developmental credits. Students were enrolled in the augmented course on the basis of identified risk and attendance at the additional sessions was made a requirement. The augmented Cell Biology course has demonstrated an improvement in success rates sufficient to convince the faculty that other courses should be similarly reformed.

In the same faculty, a first-year semester-long mathematics course has been developed into an 'extended course' that now runs over the entire year for students identified as being 'at risk' because of poor results in mathematics in the National Senior Certificate. Again, funding from the development grants has been used to pay for the additional tuition costs.

An increasing number of universities are now using 'big data' in the form of tracking mechanisms to identify areas of particular difficulty. All too often, however, and as indicated earlier, remedial measures intended to address problem areas focus on the provision of additional tutorials or other forms of intervention which are not formally curricularised. Reconsidering the curriculum structure in ways outlined above offers an alternative to this sort of approach based on sound research and theory.

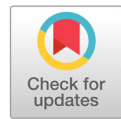
Clearly, the use of the curriculum structure to enhance teaching and learning in order to improve performance has implications for the availability of academic advising. Career advising at school is often poor or non-existent. As a result, students choose universities and programmes based on popular perceptions of what a career might involve (law being a case in point) or which do not sufficiently take account of their own interests, strengths and weaknesses. Case et al.⁹ show how participants in their study used advice provided by individuals they came across in the course of their studies in ways which often allowed them to survive and complete. Often, such advice was provided 'by chance' by an individual. This individual may have been an academic who happened to make themselves available to students through an 'open door' policy or generally friendly nature or, alternatively, someone students met who took a particular interest in them for one reason or another. Formal academic advising structures, which will allow students to make sense of their results and to plot a way forward either as they enter a university or as they progress through their studies when things are not going as expected, are an important means of enhancing performance, particularly in an institution in which attention has been paid to using the curriculum structure to improve teaching and learning.

In conclusion, therefore, curriculum restructuring offers an opportunity for those of us working in South African universities to make a real difference to our students' chances of success. The curriculum is something that students need to follow. Optional initiatives are exactly that and are opportunities for learning that are often ignored by students under pressure, an observation made by those working in academic development^{12,13} as long ago as the 1980s. While we have been presented with the opportunities to rework curricula, we have not always made the most of these. Perhaps now is the time to do so.

References



1. Scott I, Yeld N, Hendry J. A case for improving teaching and learning in South African higher education. *Higher Education Monitor* 6. Pretoria: Council on Higher Education; 2007. Available from: http://www.che.ac.za/media_and_publications/higher-education-monitor/higher-education-monitor-6-case-improving-teaching
2. Council on Higher Education. *Vital stats*. Pretoria: Council on Higher Education; 2016. Available from: http://www.che.ac.za/media_and_publications/monitoring-and-evaluation/vitalstats-public-higher-education-2016
3. Boughey C. The significance of structure, culture and agency in supporting and developing student learning at South African universities. In: Dhunpath R, Vithal R, editors. *Alternative access to higher education: Underprepared students or underprepared institutions*. Cape Town: Pearson; 2012. p. 62–88.
4. Act 58 of 1995, Republic of South Africa.
5. Kraak A. Policy ambiguity and slippage: Higher education under the new state, 1994-2001. In: Kraak A, Young M, editors. *Education in retrospect: Policy and implementation since 1990*. Cape Town: HSRC Press; 2001.
6. Ministry of Education. *The restructuring of the higher education system in South Africa*. Pretoria: Ministry of Education; 2002.
7. Wheelahan L. *Why knowledge matters in the curriculum*. London: Routledge; 2010. <https://doi.org/10.4324/9780203860236>
8. Maton K. *Knowledge and knowers: Towards a realist sociology of education*. Abingdon: Routledge; 2013. <https://doi.org/10.4324/9780203885734>
9. Case JM, Marshall D, McKenna S, Mogashana D. *Going to university: The influence of higher education on the lives of young South Africans*. Cape Town: African Minds; 2018.
10. South African Department of Higher Education and Training. *Higher education qualifications sub-framework*. *Government Gazette*. 2014;592(38116):3–42.
11. South African Department of Education. *Foundation programme grant policy*. Pretoria: Department of Education; 2004.
12. Yeld N. Credit and compulsion for ASP courses: Gains and losses. In: *Proceedings of the 1986 ASP Conference*. Pietermaritzburg: University of Natal; 1986.
13. Foggin C. A description and assessment of the recent ASP initiatives at the University of Durban Westville with particular reference to the programme in the Faculty of Arts. In: *Proceedings of the 6th Annual Conference of the South African Association for Academic Development*. Johannesburg: University of the Witwatersrand; 1991.





Searching for David within the Goliath of alien woody plant invasions in the Western Cape Province

AUTHORS:

Anthony J. Mills¹ 
Jessica L. Allen² 

AFFILIATIONS:

¹Department of Soil Science, Stellenbosch University, Stellenbosch, South Africa

²C4 EcoSolutions, Cape Town, South Africa

CORRESPONDENCE TO:

Anthony Mills

EMAIL:

mills@sun.ac.za

KEYWORDS:

Cape Floristic Region; soil nutrients; Catabolic Theory; fynbos

HOW TO CITE:

Mills AJ, Allen JL. Searching for David within the Goliath of alien woody plant invasions in the Western Cape Province. *S Afr J Sci.* 2018;114(9/10), Art. #a0285, 3 pages. <https://doi.org/10.17159/sajs.2018/a0285>

PUBLISHED:

11 September 2018

Despite a few victories in which biocontrol has contained some species of invasive alien woody plants, the war to contain alien invasions and to prevent them from threatening the exceptional plant diversity of the Cape Floristic Region is largely being lost.¹⁻⁸ Indeed, these invading plants are estimated to be expanding their range at an average rate of 7% per year.⁹

Apart from releasing more biocontrols and trying to chop out the plants where feasible, are there any other options available? Perhaps inadvertent experiments have been performed by land managers which could provide insights into new ways of managing this alien woody scourge? In densely invaded landscapes, patches of land are on occasion encountered which are mysteriously free of the alien woody invaders. Have these patches been cleared meticulously by land managers, or are there other forces at play? Perhaps the soil conditions are not suitable for the germination of the woody plants' seeds, or there are specific seed predators present in the patches, but not in the surrounding landscape? Or perhaps the plants in these patches are more competitive than the invading woody plant seedlings and prevent them from recruiting by smothering them above ground, strangling their roots below ground, releasing toxins in the soil that damage them, or taking up nutrients faster than them?

Taking these questions into account, we searched for sites in and around the southwestern Cape Floristic Region where there were sizeable patches (at least several hectares) of non-invaded land within densely invaded landscapes. We found seven such sites where there were no plausible reasons for why alien woody plants had not invaded certain patches of land. Land use histories of the sites are provided in Table 1. Three of these tree-free patches (at the Blaauwberg, Rooshook and Vergelegen sites) were old agricultural lands, with mixes of herbaceous alien weeds, alien grasses and ruderal fynbos. The other four patches (at the Bothasig, Dassenberg, Klein Dassenberg and Joostenbergkloof sites) were in previously uncultivated vegetation i.e. intact fynbos. We analysed the chemistry of topsoils in the non-invaded patches and the surrounding invaded landscape.

Table 1: Land history and vegetation type (by Mucina and Rutherford²⁵) of the seven study sites

Site	Sub-sites (n)	Land history	Vegetation type
Blaauwberg ¹	N-En (20) Mod (17) En (20)	N-En sub-sites were previously cultivated and left fallow since the late 1980s. En sub-sites were uncultivated.	Cape Flats Sand Fynbos
Bothasig ²	N-En (5) En (5)	Before 2010, the entire site was managed as a municipal park and mowed extensively. All sub-sites were uncultivated.	Cape Flats Sand Fynbos
Dassenberg ³	N-En (20) En (20)	The site was historically used for grazing. All sub-sites were uncultivated.	Atlantis Sand Fynbos
Joostenbergkloof ⁴	N-En (20) En (20)	All sub-sites were uncultivated.	Swartland Shale Renosterveld
Klein Dassenberg ⁵	N-En (18) En (20)	The site was historically used for grazing. All sub-sites were uncultivated.	Atlantis Sand Fynbos
Rooshook ⁶	N-En (18) En (21)	N-En sub-sites were fallow lands last cultivated 15 years ago. En sub-sites were uncultivated.	Swartland Shale Renosterveld and Swartland Alluvium Fynbos
Vergelegen ⁷	N-En (20) ModAg (10) ModP (10) En (20)	N-En and ModAg sub-sites were last cultivated 15 years ago. ModP sub-sites were recently felled pine plantations. En sub-sites were uncultivated.	Boland Granite Fynbos

¹Blaauwberg Nature Reserve; ²Bothasig Fynbos Nature Reserve; ³Dassenberg Coastal Catchment Partnership; ⁴Joostenbergkloof Farm; ⁵Dassenberg Coastal Catchment Partnership – Klein Dassenberg Section; ⁶Rooshook Farm; ⁷Vergelegen Wine Estate

N-En, non-encroached plots, free of alien woody plants or seedlings; En, encroached plots, densely invaded with alien woody plants; Mod, moderately encroached plots, scattered with alien woody plants; ModAg, moderately encroached and historically cultivated; ModP, moderately encroached and historically pine plantations

Although it is feasible that certain soil physical or organic chemical properties had prevented germination of the alien woody plant seeds, or certain seed predators had consumed the seeds before they germinated at our study sites, we focused our research on soil inorganic nutrients, which we consider to be a more likely explanation for the constraint of the woody plants. In particular, we tested the explanatory power of what is known as the Catabolic Theory.¹⁰⁻¹² This theory holds that short plants (namely herbs, grasses and shrubs) will outcompete tree seedlings when demand for catabolic nutrients (e.g. P, Cu and Zn) is met by supply – with this demand being

partly dependent on the availability of anabolic nutrients (e.g. B, Mg and Mn).¹³⁻¹⁵ The reasoning underpinning the theory is that if a soil nutritional regime favours the production over the metabolism of photosynthates, then a photosynthate surplus arises, which will be used for synthesising wood. The concept of short plants outcompeting tree seedlings is not new¹⁶, and the Catabolic Theory has been corroborated by studies on treeless vegetation surrounded by treed vegetation in both Australia and South Africa^{14,15}.

The results of our investigation support the Catabolic Theory. Although many nutrients are likely to be involved in the interplay between the short plants in the non-invaded patches and the alien woody invaders, two nutrients were particularly noteworthy: phosphorus (P) and boron (B). Old agricultural lands were consistently enriched in P, compared with the adjacent invaded landscapes (Figure 1). This observation suggests that an ample amount of P allows the ruderal herbaceous species to outcompete the alien woody plants. Moreover, in the case of B, all of the non-encroached sites, with the exception of Joostenbergkloof, had significantly lower concentrations of B than did the surrounding invaded landscape (Figure 2). This finding suggests that a scarcity of B favours fynbos plants over woody plant invaders.

What are the management implications of our findings? Although soil amendments are not traditional weapons in the arsenal of

conservationists, they could be used to manipulate the availability of certain nutrients, ultimately constraining the invasion of woody plants into fynbos. For example, clay minerals such as illite and vermiculite are strong binders of B.¹⁷⁻¹⁹ Based on the findings of our study, their application could potentially increase the competitive strength of fynbos plants relative to woody alien invaders. In certain farming environments, the application of P fertilisers may be appropriate for constraining woody plant invasions by increasing the competitive strength of grasses and herbs. Plot-scale experiments would be needed to determine the efficacy of such treatments before landscape interventions could be recommended. Soil amendments on a landscape-scale may not even be necessary to halt woody plant invasions – there may be opportunities to create strips of land resilient to woody plants around heavily infested landscapes. Such strips could be used to contain the infestations in much the same way that firebreaks are used to contain fires.

Although soil amendments would be a radical approach to conserving fynbos, the current threat from woody alien invaders is so grave that all options need to be put on the table and carefully scrutinised, no matter how startling. Furthermore, new ideas and new experiments are needed to develop a deeper understanding of why the woody alien invaders are so pervasive in the first place. Without such an understanding, it is difficult to pioneer new options for management in a coherent, systematic manner.

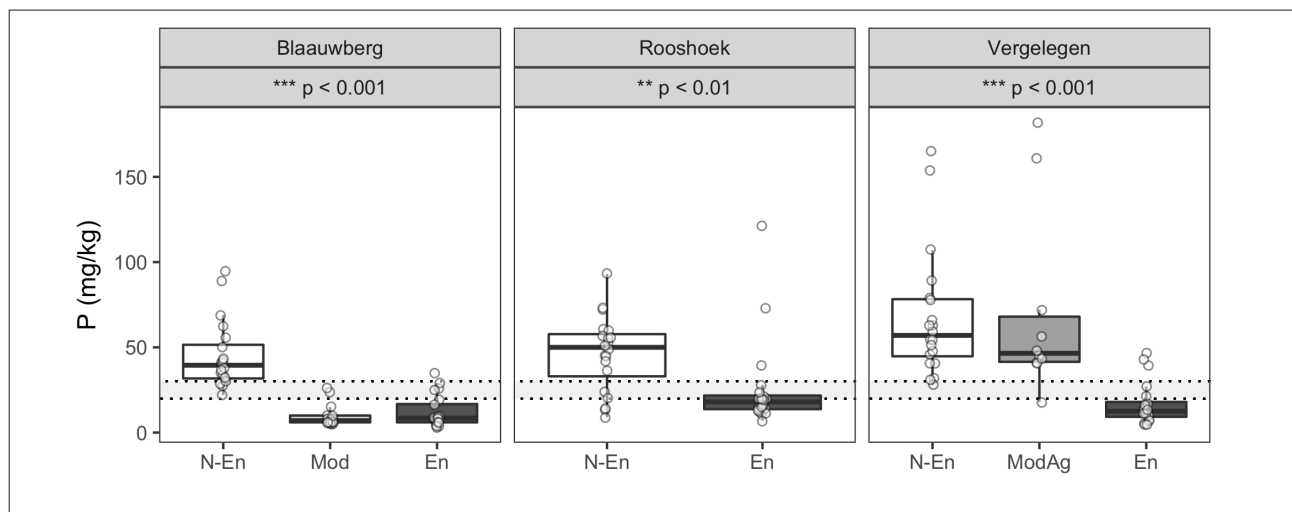


Figure 1: Phosphorus concentration (mg/kg) of encroached (En), moderately encroached (Mod), non-encroached (N-En), and moderately encroached historically cultivated (ModAg) patches at the four previously cultivated study sites. Horizontal grey bars indicate critical deficiency ranges (20–30 mg/kg).

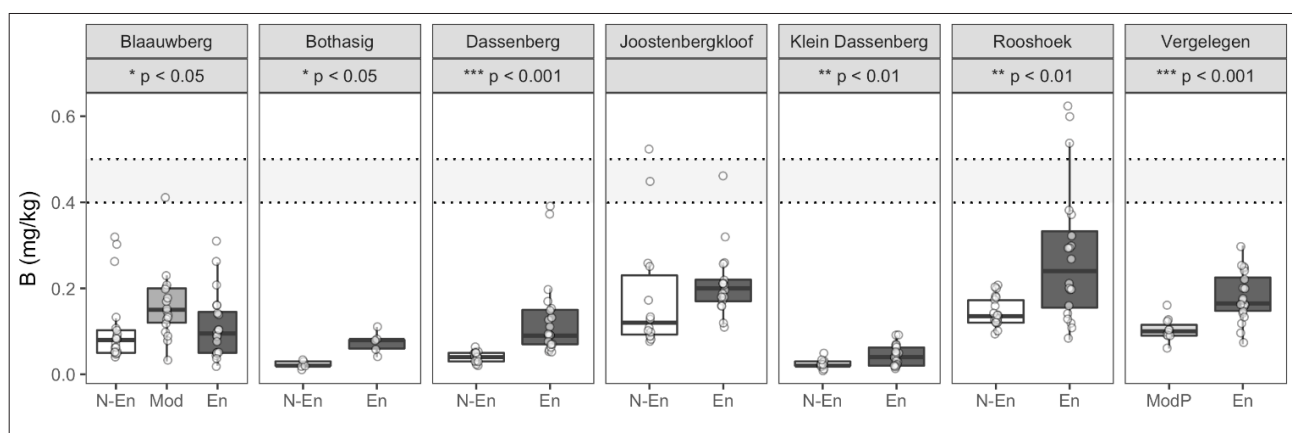


Figure 2: Boron concentration (mg/kg) of encroached (En), moderately encroached (Mod), non-encroached (N-En), and moderately encroached historically pine plantation (ModP) patches at the seven study sites. Horizontal grey bars indicate critical deficiency ranges (0.4–0.5 mg/kg). Fertilisation of agricultural lands at Vergelegen (N-En and ModAg) greatly increased B concentrations. These fertilised sites are consequently excluded from the graph.

The simple explanations of a lack of natural predators²⁰⁻²² and carbon dioxide fertilisation^{23,24} are surely just valuable chapters in a complex story. A long-term fertilisation experiment in a South African savanna has shown conclusively that changes in soil nutrient regimes can either promote or entirely constrain woody plant encroachment.¹⁵ It is highly probable that soil nutrients exert similar strong controls on woody plants in all biomes globally, not only in South African savannas. There are 13 plant nutrients which have the potential to affect the competitive interplay between fynbos plants and their woody alien invaders. It is high time that we started understanding these effects.

Acknowledgements

We thank all land owners and land managers for facilitating the research in the various nature reserves; Antoni Milewski for the discussions on concepts presented; and Adele Cormac, Zurelda M. le Roux, Selina Mochrie, Sean Foden, Ruan de Wet and M.J. Stowe for technical assistance. We also gratefully acknowledge the South African Department of Environmental Affairs' Natural Resources Management Programme and the National Research Foundation (grant number FA2005040700027) for funding this research.

References

1. Gaertner M, Richardson DM, Privett SDJ. Effects of alien plants on ecosystem structure and functioning and implications for restoration: Insights from three degraded sites in South African fynbos. *Environ Manage*. 2011;48(1):57–69. <https://doi.org/10.1007/s00267-011-9675-7>
2. Mostert E, Gaertner M, Holmes PM, Rebelo AG, Richardson DM. Impacts of invasive alien trees on threatened lowland vegetation types in the Cape Floristic Region, South Africa. *S Afr J Bot*. 2017;108:209–222. <http://dx.doi.org/10.1016/j.sajb.2016.10.014>
3. Van Rensburg J, Van Wilgen BW, Richardson DM. Reconstructing the spread of invasive alien plants on privately-owned land in the Cape Floristic Region: Vergelegen Wine Estate as a case study. *S Afr Geogr J*. 2017;6245:1–16. <http://doi.org/10.1080/03736245.2017.1340187>
4. Richardson DM, Van Wilgen BW. Invasive alien plants in South Africa: How well do we understand the ecological impacts? *S Afr J Sci*. 2004;100:45–52.
5. Gaertner M, Den Breeyen A, Hui C, Richardson DM. Impacts of alien plant invasions on species richness in Mediterranean-type ecosystems: A meta-analysis. *Prog Phys Geogr*. 2009;33(3):319–338. <https://doi.org/10.1177/0309133309341607>
6. Van Wilgen BW, Fill JM, Baard J, Cheney C, Forsyth AT, Kraaij T. Historical costs and projected future scenarios for the management of invasive alien plants in protected areas in the Cape Floristic Region. *Biol Conserv*. 2016;200:168–177. <http://dx.doi.org/10.1016/j.biocon.2016.06.008>
7. Moran VC, Hoffmann JH. Conservation of the Fynbos biome in the Cape Floral Region: The role of biological control in the management of invasive alien trees. *BioControl*. 2012;57(2):139–149. <https://doi.org/10.1007/s10526-011-9403-5>
8. Moran VC, Hoffmann JH, Zimmermann HG. Biological control of invasive alien plants in South Africa: Necessity, circumspection, and success. *Front Ecol Environ*. 2005;3:71–77. [https://doi.org/10.1890/1540-9295\(2005\)003\[0071:BCOIAIP\]2.0.CO;2](https://doi.org/10.1890/1540-9295(2005)003[0071:BCOIAIP]2.0.CO;2)
9. Rouget M, Richardson DM, Cowling RM, Lloyd JW, Lombard AT. Current patterns of habitat transformation and future threats to biodiversity in terrestrial ecosystems of the Cape Floristic Region, South Africa. *Biol Conserv*. 2003;112:63–85. [https://doi.org/10.1016/S0006-3207\(02\)00395-6](https://doi.org/10.1016/S0006-3207(02)00395-6)
10. Milewski AV, Mills AJ. Does life consistently maximise energy intensity? *Biol Rev*. 2010;85(4):859–879. <https://doi.org/10.1111/j.1469-185X.2010.00131.x>
11. Milewski AV, Mills AJ. Why was the Highveld treeless? Looking laterally to the Pampas for global edaphic principles beyond biogeographical accidents. *S Afr J Bot*. 2015;101:98–106. <https://doi.org/10.1016/j.sajb.2015.05.019>
12. Mills AJ, Milewski AV, Rogers KH, Witkowski ETF, Stalmans M. Boundary of treeless grassland in relation to nutrient content of soils on the Highveld of South Africa. *Geoderma*. 2013;200:165–171. <http://dx.doi.org/10.1016/j.geoderma.2013.02.007>
13. Mills AJ, Milewski AV, Fey MV, Gröngroft A, Petersen A, Sirami C. Constraint on woody cover in relation to nutrient content of soils in western southern Africa. *Oikos*. 2013;122(1):136–148. <https://doi.org/10.1111/j.1600-0706.2012.20417.x>
14. Mills AJ, Milewski AV, Sirami C. A preliminary test of catabolic nutrients in explanation of the puzzling treelessness of grassland in mesic Australia. *Austral Ecol*. 2016;41(8):927–937. <https://doi.org/10.1111/aec.12385>
15. Mills AJ, Milewski AV, Snyman D, Jordaan JJ. Effects of anabolic and catabolic nutrients on woody plant encroachment after long-term experimental fertilization in a South African savanna. *PLoS One*. 2017;12(6), e0179848, 24 pages. <https://doi.org/10.1371/journal.pone.0179848>
16. Knoop AWT, Walker BH. Interactions of woody and herbaceous vegetation in a southern African savanna. *J Ecol*. 1985;73(1):235–253. <https://doi.org/10.2307/2259780>
17. Hingston FJ. Reactions between boron and clays. *Soil Res*. 1964;2(1):83–95. <https://doi.org/10.1071/SR9640083>
18. Keren R, Mezuman U. Boron adsorption by clay minerals using a phenomenological equation. *Clays Clay Miner*. 1981;29(3):198–204. <https://doi.org/10.1346/CCMN.1981.0290305>
19. Kim Y, Kirkpatrick RJ. 11B NMR investigation of boron interaction with mineral surfaces: Results for boehmite, silica gel and illite. *Geochim Cosmochim Acta*. 2006;70(13):3231–3238. <https://doi.org/10.1016/j.gca.2006.04.026>
20. Andersen AN. How important is seed predation to recruitment in stable populations of long-lived perennials? *Oecologia*. 1989;81(3):310–315. <https://doi.org/10.1007/BF00377076>
21. Crawley MJ. Insect herbivores and plant population dynamics. *Annu Rev Entomol*. 1989;34(1):531–562. <https://doi.org/10.1146/annurev.en.34.010189.002531>
22. Maron JL, Vilà M. When do herbivores affect plant invasion? Evidence for the natural enemies and biotic resistance hypotheses. *Oikos*. 2001;95(3):361–373. <https://doi.org/10.1034/j.1600-0706.2001.950301.x>
23. Bradley BA, Blumenthal DM, Wilcove DS, Ziska LH. Predicting plant invasions in an era of global change. *Trends Ecol Evol*. 2010;25(5):310–318. <http://dx.doi.org/10.1016/j.tree.2009.12.003>
24. Richardson DM, Bond WJ, Dean RJ, Higgins SI, Midgley GF, Milton SJ, et al. Invasive alien species and global change: A South African perspective. In: Mooney HA, Hobbs RJ, editors. *Invasive species in a changing world*. Washington DC: Island Press; 2000. p. 303–349.
25. Mucina L, Rutherford M. *The vegetation of South Africa, Lesotho and Swaziland*. Pretoria: South African National Biodiversity Institute; 2006. p. 490–492.





The ecology of drought – a workshop report

AUTHORS:

Anthony M. Swemmer¹ 
William J. Bond^{1,2} 
Jason Donaldson³ 
Gareth P. Hempson¹ 
Johan Malherbe⁴ 
Izak P.J. Smit^{3,5} 

AFFILIATIONS:

¹South African Environmental Observation Network, Phalaborwa, South Africa

²Department of Biological Sciences, University of Cape Town, Cape Town, South Africa

³School of Animal, Plant and Environmental Sciences, University of the Witwatersrand, Johannesburg, South Africa

⁴Natural Resources and the Environment, Council for Scientific and Industrial Research, Pretoria, South Africa

⁵South African National Parks, Skukuza, South Africa

CORRESPONDENCE TO:

Anthony Swemmer

EMAIL:

tony@saeon.ac.za

KEYWORDS:

climatology; herbivory; mortality; resilience; vegetation

HOW TO CITE:

Swemmer AM, Bond WJ, Donaldson J, Hempson GP, Malherbe J, Smit IPJ. The ecology of drought – a workshop report. *S Afr J Sci.* 2018;114(9/10), Art. #5098, 3 pages. <https://doi.org/10.17159/sajs.2018/5098>

PUBLISHED:

11 September 2018

The 2014–2016 drought experienced across the summer rainfall region of South Africa brought a range of social and economic impacts, with regular coverage in the public media. However, little has been reported on the ecological impacts of the drought in the rural rangelands and protected areas of the country. To address this issue a workshop on the ecology of drought was organised by the South African Environmental Observation Network (SAEON), a facility of the National Research Foundation, in November 2017. The event brought together researchers active in savannas (bushveld) of the northeastern part of the country, including the Kruger National Park (KNP) and adjacent Lowveld, and in savannas and grasslands of KwaZulu-Natal. The land uses in these areas varied from conservation to densely populated communal rangelands.

Drought impacts on commercial farmers are relatively straightforward to assess through quantification of crop yields. Ecological impacts are less tangible, particularly as many of these may only manifest years after the drought has ceased. Furthermore, while drought impacts are invariably negative in agriculture, they may have some positive impacts in ecosystems, such as reducing herbivore numbers (thus preventing overgrazing in the long term), reducing the densities of trees (thus combating bush encroachment) and providing an opportunity for 'drought adapted' flora and fauna to thrive. The primary aim of the workshop was to assess evidence available for testing the divergent predictions of drought impacts on natural systems. A key question was how severely livestock and wildlife populations were affected. On the topic of vegetation, some argued that droughts open up savannas by killing trees, others that droughts promote the seedling recruitment of trees (as a consequence of reduced competition from grasses and less fire) leading to bush encroachment. Many questions exist around recovery following the drought. Some argued that grasses take decades to recover their productive potential, others that recovery is very rapid once the drought breaks. How does the pattern of recovery vary with the ecological status preceding droughts and actions after it? And, most ominously, does each successive drought transform an ecosystem, pushing it towards a new state from which recovery to an original state eventually becomes impossible. Alternatively, are our fauna and flora so well adapted that they can recover comfortably before the next drought strikes? This question is particularly pertinent, considering the prediction that global climate change will result in droughts becoming more severe and possibly more frequent. In addition to sharing experiences of the recent drought, we hoped to identify key data deficiencies in the various research and monitoring projects conducted during the drought, and promote the type of research and monitoring that will allow us to better capture the effects of the next major drought.

Climatology of the 2014–2016 drought

Johan Malherbe (CSIR) presented an assessment of the drought from a climatological perspective, using national-scale rainfall interpolations to map where rainfall deficits were most severe. As with major droughts of the recent past, this drought was associated with an El Niño event. It was not as severe as the droughts in the early 1960s, early 1980s and the early 1990s at a national scale. However, northeastern Mpumalanga, the eastern half of the Free State and northern KwaZulu-Natal appear to have experienced more severe drought than in the previous droughts, with Standardized Precipitation Index values of severe to extreme drought over a 2-year period (Figure 1). The large spatial variation in the severity of drought emerged as a major complication for generalising about the severity and impact of drought at a national scale. Furthermore, droughts vary in their development over time, with some areas experiencing only 1 year of below-average rainfall, and others experiencing 3 years. A key feature of the 2014–2016 drought which distinguished it from former ones was high temperatures, with an unprecedented number of locations experiencing record-breaking monthly mean maximum temperatures during 2015 and 2016. This feature reflects the global pattern of 2016 and 2015 having the highest and second highest global average temperatures on record, respectively.

Impacts on herbivores

Studies on wildlife and communal livestock populations were presented at the workshop, with die-offs being widely reported. Mortality levels varied with species, functional type and management contexts, especially those that determined the ability of animals to move in response to shortages of food and water. Izak Smit (South African National Parks) explained how management approaches in the Kruger National Park (KNP) have shifted from previous droughts, with the new philosophy promoting natural spatial variability across landscapes to provide a buffer against drought effects. This approach led to the closure of artificial waterholes and removal of fences with neighbouring protected areas since the previous severe drought. As a result, animal distributions prior to the drought were different and forage resources most likely accumulated in areas far from water sources. Large-scale culling of mega-herbivores (buffalo, hippo and elephant) ceased shortly after the 1992/1993 drought, and densities of these species were higher at the onset of the drought, compared to those for previous droughts. However, buffalo mortalities in KNP between 2015 and 2017 were 26% compared with 48% during the previous drought (between 1991 and 1993), and less than those reported in fenced private game reserves in the 1982–1983¹ drought. Carla Staver's (Yale University) analyses of rainfall and grass biomass indicated less severe drought conditions in the north of the park which, combined with larger distances between rivers and other water sources, resulted in greater grass forage reserves. Accordingly, buffalo appear to have migrated from the south of the park to make use of these reserves. In contrast to buffalo, hippo populations in KNP suffered heavy mortality (45% across the four major perennial rivers of the park) with these water-bound animals unable to access areas of forage reserves once they had depleted all forage within walking distance of the large perennial rivers.

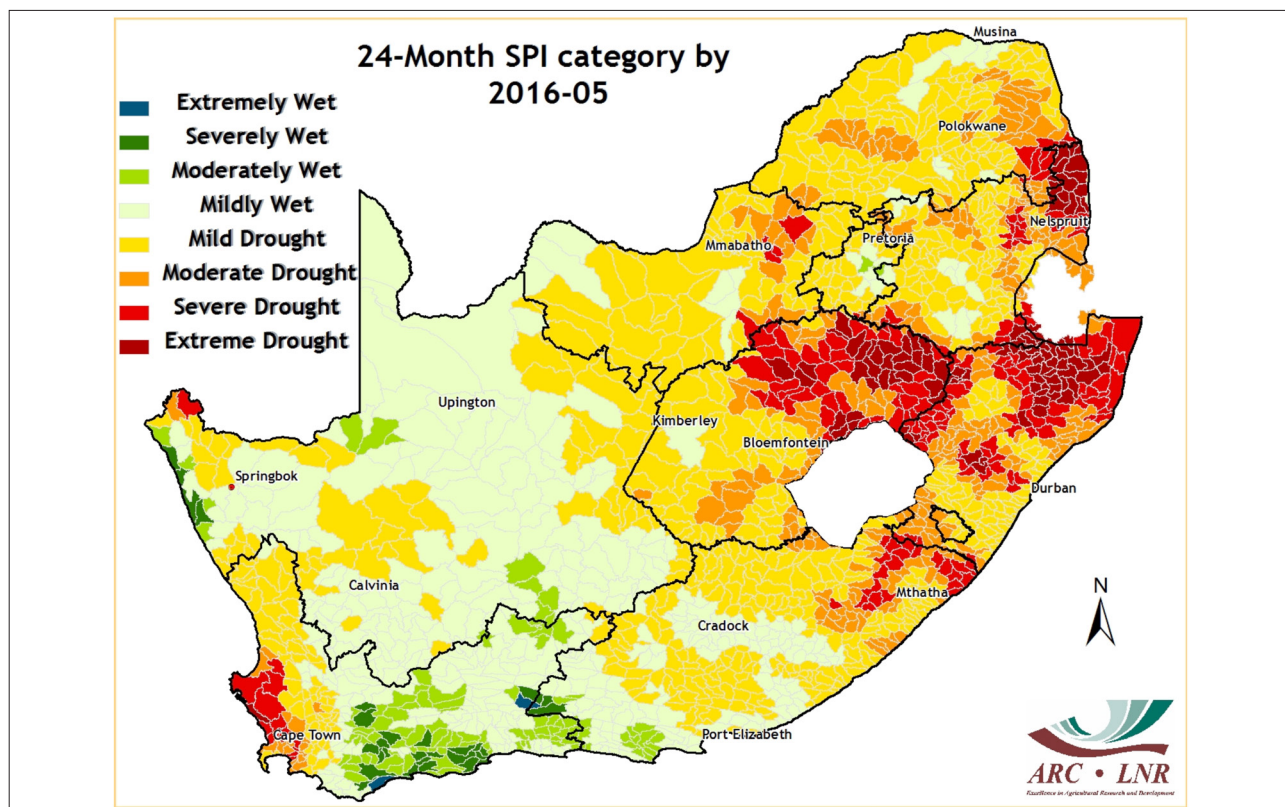


Figure 1: The Standardised Precipitation Index for South Africa over a 2-year period, as of May 2016, indicating which parts of the country experienced the most severe rainfall deficits.

Mike Peel (Agricultural Research Council) highlighted the importance of herbivores being able to move freely during droughts, reporting over 75% mortality of herbivores in a small, fenced game reserve. Neighbouring private reserves that were open to KNP did not experience the same level of drought mortalities. Researchers working in the Lowveld identified an additional factor that may have mitigated herbivore mortality during the 2014–2016 drought: the timing of rainfall. Rainfall events that occurred in the late summer and early winter produced pulses in grass growth, extending the availability of green forage. Browsing herbivores were seemingly less affected than grazers by the drought. Against expectations, elephant numbers in KNP increased by 13% between 2015 and 2017. This suggests that the elephant population is still far from being resource limited, and prompted the question of whether drought impacts on the tree layer might have been more pronounced if elephant densities were two or three times higher.

The reliability of livestock data from communal rangelands is problematic because extension officers do not always record livestock numbers accurately, and some livestock owners no longer make use of dipping tanks (where livestock numbers are recorded) nor report stock losses. Estimates of mortality need to take these uncertainties and different methodologies into account. Teddy Mnisi (Department of Agriculture, Limpopo Province) presented the results of monthly monitoring of livestock per municipal district, showing extensive losses and cattle mortality as high as 33% in the Giyani District. Cobus Botha (Department of Agriculture and Rural Development, KwaZulu-Natal) presented mortality numbers for northern KwaZulu-Natal, for which losses were extensive. Rauri Alcock (Mdukatshani Rural Development Programme) revealed even higher mortality rates from an intensive study of livestock production in KwaZulu-Natal, with losses of 43% for cattle and 29% for goats. Losses were heaviest for owners with smaller herds of both cattle and goats, and goat numbers recovered more rapidly than cattle numbers in the year following the drought. Using household-level data, Rauri estimated that the economic losses from livestock across KwaZulu-Natal were more than ZAR10 billion. This amount far exceeds

the level of drought relief that government is able to provide. In any case the provision of supplementary feed for cattle, at such a large scale, is not logistically feasible and would result in further degradation of communal rangelands by supporting livestock densities well in excess of local carrying capacities.²

Impacts on vegetation

At all sites where the grass layer was studied, grass production was severely reduced (as expected) but changes in species composition and recovery of production following the drought showed some interesting and unexpected patterns. Many researchers working in the Lowveld identified late season rainfall as a key factor that prevented even more severe reductions in production, and most likely enabled many herbivores to survive the subsequent winter. Longer-term impacts were revealed from monitoring data in conservation areas and commercial livestock farms. These impacts included a reduction in the basal cover of perennial grasses, at both mesic and semi-arid sites, with increases in the abundance of forbs at drier sites, as has been observed for previous droughts.³ Dave Thompson (SAEON) presented results from a long-term research site near Satara (central KNP) where drought was particularly severe and where grazing is heavy (relative to KNP in general). Here, grass production was surprisingly resilient, returning to near pre-drought levels in 2016–2017. Moreover, grass composition showed a distinct change towards more palatable species, with large declines in unpalatable perennials, and an increased abundance of species of high-grazing value (both perennial and annual). A similar trend was presented for long-term research plots in Hluhluwe-iMfolozi Park (KwaZulu-Natal) by Huyam Altayeb (University of Cape Town). These positive results may be offset by greater erosion risk in areas where established unpalatable perennials were lost, although a paucity of data on basal cover and erosion rates makes this difficult to assess. However, it was clear that recovery of primary productivity can occur over years rather than decades, at least in relatively pristine ecosystems.

Extensive tree deaths from drought have been reported for savannas in the USA and Australia⁴⁻⁶, but previous droughts in South Africa appear to have had less impact⁷⁻⁸. Less than 1% of trees had died by the end of 2016 in Hluhluwe-iMfolozi Park, but there was some significant mortality in southern KNP. Research presented by Carla Staver (Yale University) and Tony Swemmer (SAEON) showed that mortality of trees and shrubs was as high as 40% at some sites within KNP, and even appeared to have resulted in the conversion of open savanna to grassland at a few of these. However such impacts were not widespread, and mortalities only occurred in certain regions and only affected certain species. Overall, it was clear that the drought had not caused widespread tree or shrub mortality across the savanna biome. Whether the drought will result in a pulse of tree recruitment remains to be seen, and established long-term research at a number of sites in the Lowveld provides an opportunity to address this.

Issues emerging

A number of key research gaps emerged from discussions. It was clear that annual rainfall is not a sufficient metric to properly understand, and predict, the ecological impacts of drought. Other metrics, such as the length of intervals between rainfall events or the incidence of rain in the dry season, also need to be considered, as well as the effect of higher temperatures. Late summer and early winter rainfalls may be particularly relevant for determining how long green forage and browse persist as the dry season progresses, and thus may be a key determinant of herbivore mortality. A better understanding of how El Niño events translate into changes in these variables, at regional scales, is needed to predict the impacts of projected changes in the frequency and magnitude of El Niño events. El Niño development at a monthly timescale may be important in this regard. Spatial variation in rainfall also needs to be measured adequately to properly understand the severity of drought at national or regional scales.

The reliability of the existing data on livestock numbers, especially in communal areas, needs to be assessed. These data are the only means of measuring drought impacts on herbivores over vast areas of the country, as well as quantifying the economic impacts of droughts. Better information on the movement patterns of herbivores (domestic and wild) would enhance our understanding of how different species respond to drought conditions, and enable us to identify forage reserves that are crucial for sustaining populations.² Little is known regarding the potential importance of disease in exacerbating drought-induced mortality of herbivores. Finally, little is known regarding the consequences for rural poor when livestock herds die.

Regarding vegetation data currently available for studying drought impacts, there is a strong geographical bias. There is apparently little data for many of the areas that were severely affected by the 2014–2016 drought, particularly in the eastern Free State, North West and the Northern Cape Provinces. Most studies of the herbaceous layer presented at the workshop were from semi-arid areas with fertile soils or high rainfall. These are areas that support high herbivore densities, and are thereof of most interest to those concerned about the impact of the drought on herbivore productivity. However, semi-arid areas with infertile soils, such as those underlain by granitic rocks, are far more extensive and make a major contribution to livestock production and eco-tourism at the national scale. Wider geographical coverage of veld and herbivore monitoring is needed before we can generalise drought effects countrywide. Remote sensing of drought impacts could potentially be very useful in extending vegetation monitoring but a key limitation is that dead shrubs and smaller trees cannot be identified, limiting large-scale assessments of mortality. Finally, the mechanisms through which drought kills trees was identified as an important research question for predicting the impacts of more severe droughts in the future.

Trees can starve to death if they are unable to store sufficient carbon as a result of insufficient carbon gain over multiple summers of low rainfall. Alternatively, mortality may be caused by embolisms in the xylem which occur when soil moisture drops below a critical level, in which case a single year (or even a few months) of severe drought could be all that is required. Another gap of 'drought research' identified was whether drought results in a pulse of tree recruitment once good rains return, and studies on post-drought seedling recruitment were agreed to be important. There were no studies presented, or known about by workshop participants, regarding the question of whether drought promotes the spread of alien plants, which Maanda Lighava (University of Venda) raised as a key concern for rural rangelands in the northern Lowveld. Finally, the consequences of past and future vegetation change on stream flow, and thus the supply of fresh water during a drought, is a highly relevant research frontier for South Africa.

Conclusions

The overall conclusion from the workshop was that the 2014–2016 drought did not have catastrophic impacts on the ecosystems of protected areas and rural rangelands in the northern and eastern parts of the country. Herbivore declines were only severe for a few species, and more so in heavily stocked rangelands and game farms. Extensive protected areas, such as the Kruger National Park, experienced relatively lower herbivore mortality than in previous droughts, and elephants were barely affected. Livestock losses were relatively more severe for owners of small herds, and goat herds appear to have been more resilient than cattle. However, much about the long-term impacts of the drought remain unknown, particularly in terms of long-term changes to vegetation, while links between such changes and negative socio-economic impacts have not been well studied.

Acknowledgements

The workshop was hosted by SAEON, with funding from the Department of Science and Technology and the National Research Foundation. Workshop participants paid for their own travel.

References

1. Walker BH, Emslie RH, Owen-Smith N, Scholes RJ. To cull or not to cull: Lessons from a southern African drought. *J Appl Ecol*. 1987;24:381–401. <https://doi.org/10.2307/2403882>
2. Illius AW, O'Connor TG. On the relevance of nonequilibrium concepts to arid and semiarid grazing systems. *Ecol Appl*. 1999;9:798–813. [https://doi.org/10.1890/1051-0761\(1999\)009\[0798:OTRONC\]2.0.CO;2](https://doi.org/10.1890/1051-0761(1999)009[0798:OTRONC]2.0.CO;2)
3. O'Connor TG. Influence of rainfall and grazing on the compositional change of the herbaceous layer of a sandveld savanna. *J Grassl Soc South Afr*. 1991;8:103–109. <https://doi.org/10.1080/02566702.1991.9648273>
4. Carter MG. Effects of drought on mesquite. *J Range Manage Archives*. 1964;17:275–276. <https://doi.org/10.2307/3895453>
5. Fensham RJ, Holman JE. Temporal and spatial patterns in drought-related tree dieback in Australian savanna. *J Appl Ecol*. 1999;36:1035–1050. <https://doi.org/10.1046/j.1365-2664.1999.00460.x>
6. Breshears DD, Myers OB, Meyer CW, Barnes FJ, Zou CB, Allen CD, et al. Tree die-off in response to global change-type drought: Mortality insights from a decade of plant water potential measurements. *Front Ecol Environ*. 2009;7:185–189. <https://doi.org/10.1890/080016>
7. Gertenbach WPD. Rainfall patterns in the Kruger National Park. *Koedoe*. 1980;23:35–43. <https://doi.org/10.4102/koedoe.v23i1.634>
8. O'Connor TG. Impact of sustained drought on a semi-arid *Colophospermum mopane* savanna. *Afr J Range Forage Sci*. 1998;15:83–91. <https://doi.org/10.1080/10220119.1998.9647948>





Publication incentives based on journal rankings disadvantage local publications

AUTHORS:

Alan T.K. Lee¹
Carol A. Simon²

AFFILIATIONS:

¹Editor of *Ostrich: Journal of African Ornithology*, Cape Town, South Africa

²Co-editor of *African Zoology*, Stellenbosch, South Africa

CORRESPONDENCE TO:

Alan Lee

EMAIL:

alan.tk.lee@googlemail.com

KEYWORDS:

Scimago; impact factor; society journal; publication subsidy; South Africa

HOW TO CITE:

Lee ATK, Simon CA. Publication incentives based on journal rankings disadvantage local publications. *S Afr J Sci*. 2018;114(9/10), Art. #a0289, 3 pages. <https://doi.org/10.17159/sajs.2018/a0289>

PUBLISHED:

11 September 2018

The effect of globalisation on knowledge exchange, which is mediated very largely through scientific journals published in English that have their origins in Europe and North America, has resulted in the neglect of regional journals.¹ Analyses of publication output and citation impact of about 200 South African journals showed that the vast majority of local journals were virtually invisible within the global science arena.² These and related concerns have resulted in various nationally led incentive schemes to promote South African research.

The government's incentive system works through the Department of Higher Education and Training (DHET) by funding universities for articles published in accredited journals or other accepted publication outlets (books etc). The DHET does not currently differentiate between national and international publications and will award the subsidy if the publication appears on one of its accreditation lists for journal articles. Research incentive systems are believed to be one of the drivers of the increase in paper publication numbers in South Africa.

Institutions may decide how the incentive funds are spent. Some use the money for general research funds, while others give a proportion to the faculty. In the latter case, funds may be further divided between the faculty and the author. Several universities have adopted the DHET policy and procedures for measuring research output and rating and rewarding researchers for publishing papers, although there is considerable variation in how incentive funding is spent within institutions. North-West University offers larger rewards for articles in internationally published journals than local journals, while University of Cape Town offers no direct financial incentives to academic staff to publish in international or local journals. Stellenbosch University offers ~10% of the subsidy to the authors, with no distinction between journals. At the University of Johannesburg, a minimum of 70% is paid to the researcher for publication in international journals while a maximum of 30% accrues by the faculty, and in the case of DHET-accredited South African journal articles, of the subsidy transferred to the faculty, a minimum of 50% goes to the researcher and a maximum of 50% to the faculty. While most other universities pay a flat fee for publication, all actively encourage academic staff to publish in international journals because of NRF scoring criteria and, by extension, discourage them from publishing in national journals that are frequently also society journals.

This Commentary was prompted by learning that the University of KwaZulu-Natal (UKZN) will be implementing graded financial incentives to authors based on the Scimago Journal Rank (SJR) indicator quartile system. SJR is a measure of the scientific influence of scholarly journals that accounts for both the number of citations received by a journal and the importance, or prestige, of the journals from which such citations come: higher SJR values are meant to indicate greater journal prestige. Journals are then categorised into quartiles depending on this index and academic category, with journals in Q1 and Q2 associated with greater prestige (and, in the case of UKZN, higher financial incentives) than those in Q3 and Q4.

The SJR indicator is a variant of the eigenvector centrality measure used in network theory. Such measures establish the importance of a node in a network based on the principle that connections to high-scoring nodes contribute more to the score of the node. SJR is now considered an alternative to the well-established journal impact factor (IF), because of its open-access nature, larger source database, and assessment of the quality of citations.³ Our issue is with the definition of prestige, as algorithms have a particularly problematic history of processing information about race.⁴

To investigate this concern we examined the database of rated journals available at <https://www.scimagojr.com/journalrank.php?out=xls>. We selected only journals ($n=23\ 226$). An examination of SJR as a function of IF (listed as citations per document) using a negative binomial generalised linear model (even the log-transformed distribution of SJR displayed this distribution) showed a very strong effect (an increase of $\log(\text{SJR}+1)$ of 0.12 ± 0.001 per IF score, $z=84.66$, $p<0.001$). However, there is a large amount of spread in the residuals, partly a consequence of very high SJR scores for only a handful of journals.

A straightforward analysis of South African local journals was not forthcoming using the ScimagoJR database, as only 82 journals are attributed to South Africa: several local journals hosted in partnership with international publishers are listed as international, for example Taylor and Francis co-hosts *Ostrich* and *African Zoology*, the journals of BirdLife South Africa and the Zoological Society of Southern Africa, respectively. Nevertheless, SJR scores are negatively weighted for this set of journals (-0.62 ± 0.24 , $z=-2.62$, $p=0.009$). Of greater concern, for a subset of journals for which the title contained the word 'Africa' or its derivatives ($n=162$), SJR was also negatively weighted (-0.53 ± 0.16 , $z=-3.32$, $p<0.001$). By contrast, journals that contained 'America' or derivatives in their title ($n=434$) had higher SJR scores (median of $\log(\text{SJR}+1)$ for America = 0.42; Other = 0.27; $p<0.001$, Figure 1). The comparatively low scores for African journals may not only be a consequence of low visibility, but also perceptions of quality (see below).

Our concern is thus that financially weighted national incentive schemes based on SJR will further disadvantage locally published journals by effectively leaving local journals with the 'leftovers'⁵ and encouraging an institutionalised acceptance of intellectual colonialism, i.e. academics are encouraged to publish local research in high impact factor journals based overseas. In cases of author-pays models of open access (e.g. *PLoS One*), South Africa incurs a double whammy of not only exporting our research, but of paying dearly for the privilege to do so.

The importance of local journals cannot be overstated. Local research and management organisations, and even government, are usually aware of their existence and can easily keep tabs on articles they publish. By contrast, articles published overseas can be lost in what amounts to academic point scoring. As a real example, a BirdLife

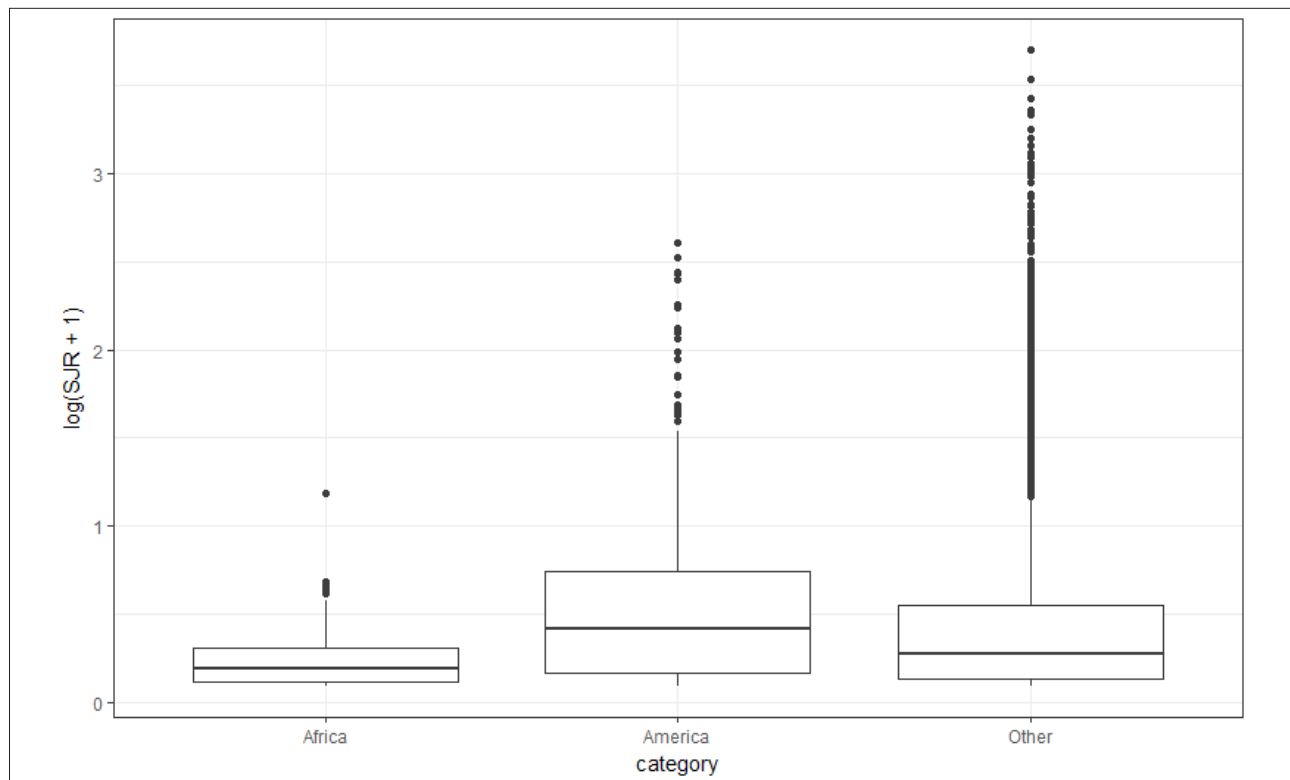


Figure 1: Box plots of log-transformed SJR scores for different journal categories: those containing 'Africa' in the title; those containing 'America' in the title, and the group of other journals. Journals with 'African American' in their titles ($n=4$) are included in the 'Africa' category.

South Africa regional division had recently embarked on a monitoring exercise to determine vulnerability of a target set of birds, only to discover later that similar information was already published abroad.⁶ This information had not been brought to the attention of the local organisation, and this certainly cost the organisation time, if not money.

Similarly, focused journals should allow readers and researchers to understand themes, trends, and knowledge gaps. This is far more difficult to undertake when research is scattered to the diaspora of available international journals. Again, to illustrate this point, the conclusions of a recently published article, entitled 'Trends and themes in African ornithology'⁷, were based on a meta-analysis of five journals which were identified to have an African ornithological focus. This article was criticised by some who pointed out that South Africa's top ornithologists were not represented in the review as they publish in ad-hoc high impact factor journals.⁸ In his response, Beale⁹ raises concern regarding 'scientific colonialism' and the lack of on-the-ground support for local researchers – sentiments echoed by Cresswell¹⁰.

The local scientific community is further harmed by the reluctance of top local researchers to publish locally when the journal in question is published by a scientific society. The main objectives of scientific societies include promoting, facilitating and encouraging research within a discipline¹¹ (e.g. ZSSA constitution, 2017). Thus, income generated from society journals is usually ploughed back into the research community.¹¹ This scenario is clearly illustrated when considering *African Zoology* – a journal which generates most of the income for its parent society, the Zoological Society of Southern Africa. This income is used to provide each society member with a subscription to the journal, seed money to organisers of the biannual conference and financial support for students to attend these, and sometimes international, conferences. In addition, the best undergraduate final-year and honours students in zoology, as nominated by their institutions, are awarded free membership to the society for a year. Income for the society, and consequently the services that it may provide for the community, increases with an increase in the number of quality articles published in their journal. For example, one of

the most highly cited and viewed papers in *African Zoology* is a review paper published in 2011. To date, it has been cited 90 times, and, in 2017 alone, it received 1008 hits on BioOne, which generated income equivalent to the cost of the prizes of free membership for undergraduate and honours students for up to 2 years.

Impact factor and similar indices are important as indicators of the quality and reach of journals, but conditions leading to initial journal IF may have changed, and instead now persist as a result of perception rather than quality: in effect, an impact factor trap. Previously, many journals published in Africa and other developing regions were not very visible in developed countries because they were not indexed in the bibliographic databases that are largely produced in developed countries.^{2,5,12} Historically, international readers struggled to access journals published in Africa and other developing countries,^{5,12} meaning that even high quality research would not be as widely read and cited as it deserved. Furthermore, society journals (which account for several local journals) publish information that may be of high quality, but not with high immediate impact, thus reducing the contribution that citations of these publications may have on the journal impact factor.¹¹ These factors would have contributed to the low impact factors achieved by such journals. As more local researchers published in international (glam) journals, local journals were essentially left with the 'leftovers',⁵ making it increasingly difficult for local journals to attain high impact factors. However, many local journals are now published by, or in partnership with, international publishing houses, or are open access, so limited access to regional journals is no longer relevant. Initiatives such as African Journals Online promote easy access to African content.¹³ In a review of the trends of 17 South African journals ranked by Thomson Reuters (now Clarivate Analytics) Journal Citation Reports®, 11 of these improved their impact factors between 2002 and 2009, albeit not enough to change quartiles for the most case.¹⁴ The point is, the factors that may originally have contributed to the low impact factors of local journals are not the factors that are now discouraging authors from submitting high quality research to them. Thus, local journals are still faced with a difficult task (attracting research away from glam journals), now made

impossible with financial incentives to authors scored only on impact factor or SJR. It is certainly already the case that at a department of the University of Pretoria which implements an economic incentive scheme to authors scaled by IF, some researchers there are reluctant to publish in *Ostrich*, despite requests to do so and even though *Ostrich*'s impact factor has been increasing over recent years.

We realise that scaled financial schemes are being considered given concerns that the DHET research output subsidy model may inadvertently penalise high-citation publication¹⁵, and that the current subsidy model has led to what has been called 'an overabundance of weak journals'¹⁶. African journals need to be given more time to play catch up with international journals before institutional graded incentives are introduced. Should academic institutions wish to pursue graded financial incentives for authors, we advocate that 'local' needs be recognised. This could easily be done by up ranking current quantile categories (e.g. a local Q2 journal could be Q3), or using an alternative ranking system e.g. altmetrics¹⁷.

We finish by reminding readers of the first recommendation from the ASSAf 2006 report¹:

Recommendation no. 1: that all stakeholders in the South African research enterprise should each in their own way support local/national research journals that actively seek to be of international quality and are indexed in an internationally recognised, bibliometrically accessible database, through following best practice in editorial discernment and peer review, including adaptations.

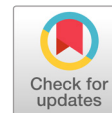
Acknowledgements

We thank the following people for assisting or commenting on draft versions of this Commentary: Ian Glenn (UFS), Derek Engelbrecht (UL), Bettine van Vuuren (UJ), Andre Ganswindt (UP), Michael Cherry (SU), Colleen Downs (UKZN) and Mike Schramm (NISC). We thank Marie Theron (SU) for sourcing literature.

References

1. Academy of Science of South Africa (ASSAf). Report on a strategic approach to research publishing in South Africa. Pretoria: ASSAf; 2006. Available from: <https://www.assaf.org.za/files/2011/02/2466-ASSAf-Strategic-approach-to-research-publishing-2.pdf>
2. Tijssen RJ, Mouton J, Van Leeuwen TN, Boshoff N. How relevant are local scholarly journals in global science? A case study of South Africa. *Res Eval*. 2006;15(3):163–174. <https://doi.org/10.3152/147154406781775904>
3. Falagas ME, Kouranos VD, Arencibia-Jorge R, Karageorgopoulos DE. Comparison of SCImago journal rank indicator with journal impact factor. *FASEB J*. 2008;22(8):2623–2628. <https://doi.org/10.1096/fj.08-107938>
4. Sandvig C, Hamilton K, Karahalios K, Langbert C. Automation, algorithms, and politics | when the algorithm itself is a racist: Diagnosing ethical harm in the basic components of software. *Int J Commun*. 2016;10:19.
5. Gibbs WW. Lost science in the third world. *Sci Am*. 1995;273(2):92–99. <https://doi.org/10.1038/scientificamerican0895-92>
6. Barshep Y, Erni B, Underhill LG, Altwegg R. Identifying ecological and life-history drivers of population dynamics of wetland birds in South Africa. *Glob Ecol Conserv*. 2017;12:96–107. <https://doi.org/10.1016/j.gecco.2017.09.001>
7. Beale CM. Trends and themes in African ornithology. *Ostrich*. 2018;89(2):99–108. <https://doi.org/10.2989/00306525.2017.1407834>
8. McKechnie AE, Amar A. Missing the bigger picture: A response to Beale (2018). *Ostrich*. 2018;89(2):151–152. <https://doi.org/10.2989/00306525.2018.1467979>
9. Beale CM. Missing the bigger picture: Reply to McKechnie and Amar (2018). *Ostrich*. 2018;89(2):153–154. <https://doi.org/10.2989/00306525.2018.1471646>
10. Cresswell W. The continuing lack of ornithological research capacity in almost all of West Africa. *Ostrich*. 2018;89(2):123–129. <https://doi.org/10.2989/00306525.2017.1388301>
11. Schloss PD, Johnston M, Casadevall A. Support science by publishing in scientific society journals. *Am Soc Microbiol*. 2017;8(5):e01633–e01637. <https://doi.org/10.1128/mBio.01633-17>
12. Lor PJ, Britz J. Knowledge production from an African perspective: International information flows and intellectual property. *Int Inf Lib Rev*. 2005;37(2):61–76.
13. Smart P. Increasing the visibility of published research: African journals online. *Afr Today*. 2005:39–53. <https://doi.org/10.1353/at.2006.0018>
14. Pouris AE, Pouris A. An assessment of South Africa's research journals: Impact factors, Eigenfactors and structure of editorial boards. *S Afr J Sci*. 2015;111(3-4), Art. #2013-0358, 8 pages. <https://doi.org/10.17159/sajs.2015/20130358>
15. Harley YX, Huysamen E, Hlungwani C, Douglas T. Does the DHET research output subsidy model penalise high-citation publication? A case study. *S Afr J Sci*. 2016;112(5–6), Art. #2015-0352, 3 pages. <https://doi.org/10.17159/sajs.2016/20150352>
16. Koenig R. A call to improve South Africa's journals. *Science*. 2006;312:831. <https://doi.org/10.1126/science.312.5775.831b>
17. Onyancha OB. Altmetrics of South African journals: Implications for scholarly impact of South African research. *Pub Res Q*. 2017;33(1):71–91. <https://doi.org/10.1007/s12109-016-9485-0>





Does South African research output promote innovation?

AUTHORS:

Roula Inglesi-Lotz¹

Anastassios Pouris²

AFFILIATIONS:

¹Department of Economics, University of Pretoria, Pretoria, South Africa

²Institute for Technological Innovation, University of Pretoria, Pretoria, South Africa

CORRESPONDENCE TO:

Roula Inglesi-Lotz

EMAIL:

roula.inglesi-lotz@up.ac.za

KEYWORDS:

R&D; publications; patents; technology

HOW TO CITE:

Inglesi-Lotz R, Pouris A. Does South African research output promote innovation? *S Afr J Sci.* 2018;114(9/10), Art. #a0286, 3 pages. <https://doi.org/10.17159/sajs.2018/a0286>

PUBLISHED:

11 September 2018

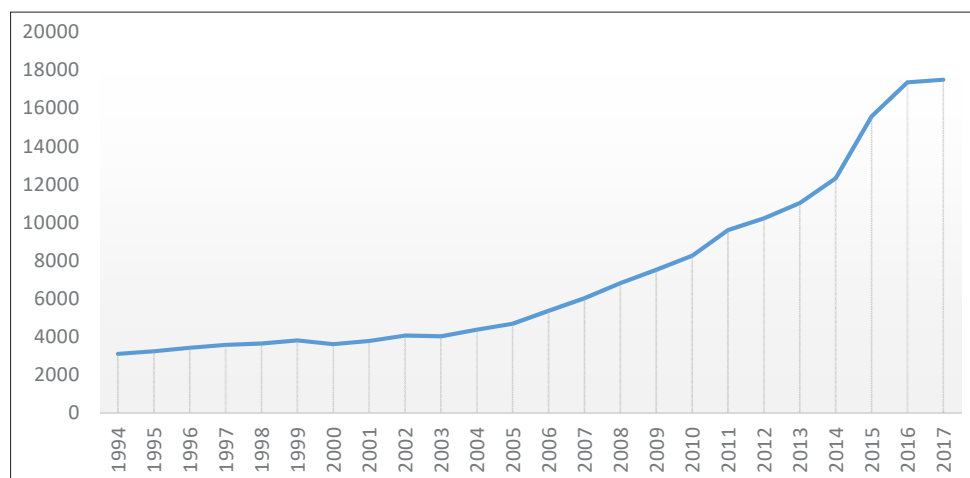
Research and development (R&D) activities that lead to innovation can contribute highly to a country's economic growth and development. However, research output can be expressed in different forms, such as research publications and patents. Patents are considered to be closer to implemented innovation than are research publications, but research publications improve human capital and propose innovative solutions that can lead to patenting activities. So although they are two different outcomes, publications and patents can work in synergy.

A recently published article entitled 'Patents vs publications and R&D: Three sides of the same coin? Panel Smooth Transition Regression (PSTR) for OECD and BRICS countries'¹ in the journal *Applied Economics*, identified the existence of research thresholds beyond which countries can move in the next stage and increase their number of patents. The concept tested was that a minimum number of publications (as a proxy to research capacity) should be reached before a country or a region can engage in patenting activities that will eventually promote higher innovation and consequently economic growth.

In that paper, we investigated the existence of thresholds in the OECD and the BRICS countries and found that the threshold of publications before patenting activity started was 11 113 for the whole sample, 9038 for the OECD countries and 9945 for the BRICS countries.

The effect of the number of publications and R&D as variables of transition is opposite. Beyond the threshold, the number of publications increases the possibility of patenting for the two groups of countries. However, beyond the threshold of R&D this variable decreases the probability of patenting. From these results and based on the optimal threshold, we conclude that patents in OECD and BRICS countries are positively influenced by the number of publications.¹

Based on these thresholds, South Africa has not yet reached the relevant level of the estimated threshold even though the country's research publications (in all disciplines) have increased substantially over the recent decade (Figure 1).²



Source: Clarivate Analytics²

Figure 1: South African research publications in all disciplines, 1994–2017.

The question then arises as to why South Africa has not yet reached the threshold and whether the disciplines of the publications in the country play a role. Obviously, patents require particular skills that may not be available in a country because of policy preferences and directions. This Commentary provides evidence of the current distribution of the research publications of South Africa according to five major clusters of research. The article is also important for the setting of research priorities in the country.

Empirical analysis

Globally, countries with high innovation levels and improvements in technology have generally produced more research publications in disciplines of technology than other types of science. That is to say, those countries that drive technology research globally tend to enjoy higher levels of economic growth, living standards and development.² Figure 2 presents countries' shares of the world's total research in the discipline of technology, as a representation of their strength in technology research. 'Technology' as a discipline includes: Acoustics, Automation and Control Systems; Computer Sciences; Construction & Building Technology; Energy & Fuels; Engineering; Imaging Science & Photographic Technology; Information Science & Library Science; Instruments & Instrumentation; Materials Science; Mechanics; Metallurgy & Metallurgical Engineering; Microscopy;

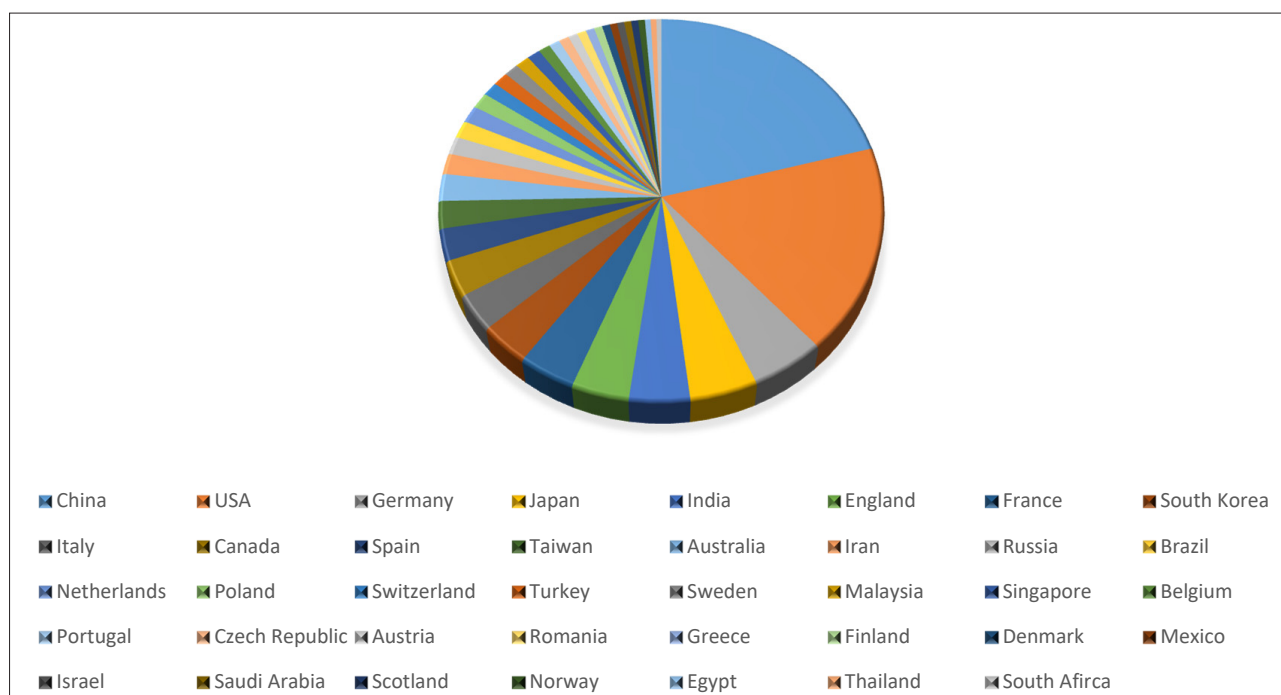
Nuclear Science & Technology; Operations Research & Management Science; Remote Sensing; Robotics; Science & Technology Other Topics; Spectroscopy; Telecommunications; Transportation.

The distribution in Figure 2 is skewed, with China and the USA producing more than 40% of the total (23.1% and 19.5%, respectively) and all the other countries producing less than 5%. Not surprisingly, developed countries such as the USA, Germany, France and Italy are in the top 10, but the BRICS countries are also well represented. China is in first place (with 23.1%) and India in fifth place (with 4.6%), while Russia is 15th with 1.7% and Brazil 16th with 1.6% of the total share in world research in the Technology field. South Africa, however, does not follow the high research productivity of the rest of the emerging technologies, producing only 0.37% of the world's share of research publications in Technology.

But is South Africa lacking in research publications in only Technology? Table 1 shows the broader scientific disciplines in which South Africa has published research outputs for the period from 1995 to 2016. The information included is the number of publications produced with at least one South African author in each broad scientific field; the share of each discipline in the South African set of publications; the country's share of the world's publications in the particular field and South Africa's world ranking in each field.

South Africa is ranked 18th in the world in Social Sciences and Arts and Humanities; 33th in Life Sciences; 38th in Physical Sciences and 40th in Technology (out of 198 countries in the data set). It is important to note that different metrics could lead to different conclusions about the country's strengths and weaknesses. For example, proponents of arts and humanities or researchers with limited information may use the number of publications in order to provide additional support in the field. South Africa's ranking in the world is probably the most revealing one as it takes into account international preferences and emphases.

From Table 1, it can be seen that South Africa's research capacity varies among the different broad research disciplines. Looking only at the number of publications in each discipline, one can observe that, for example, the Arts and Humanities discipline publishes approximately one fourth of the papers of the Technology or Physical Sciences. That observation might mislead policy into promoting R&D funding towards arts and humanities to increase their productivity. However, by evaluating the rankings, a better understanding of the world's research in this discipline is offered. A good example is the comparison between Social Sciences and Technology – at 31 thousand and 37 thousand papers, respectively – for which South Africa is ranked 18th in the former and 40th in the latter!



Source: Clarivate Analytics²

Figure 2: Research in the Technology field (country's share of the world's total, %) in 2016.

Table 1: Research performance of South Africa per discipline, 1995–2016

Discipline	Number of publications	Share of South Africa's total	South Africa's share of the world total	South Africa's world ranking (out of 198 countries)
Social Sciences	31 651	13.9%	0.80%	18
Arts and Humanities	8 842	3.9%	0.33%	18
Life Sciences	103 441	45.4 %	0.63%	33
Physical Sciences	46 119	20.2%	0.50%	38
Technology	37 847	16.6%	0.37%	40

Source: Authors' calculations on data from NACIP³

Discussion

This Commentary aims to provide a broad picture of the revealed priorities (priorities revealed by the volume of publications in particular research domains) in South Africa and how the choice of indicators may affect the relevant conclusions.

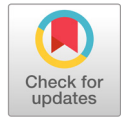
Table 1 shows that if the criterion was number of publications or share of the country's total number of publications, the verdict would have been that South Africa emphasises life sciences and neglects arts and humanities. If the criterion was South Africa's share of the world's publications, the results would have shown that South Africa makes the greatest contribution in the social sciences (0.80%). Finally, if the criterion is South Africa's ranking in the world, the results would show that the country ranks 18th in the world in social sciences and humanities and 40th in the domain of technology. The reasoning behind these variations is that different scientific publications have different publication patterns that have to be remedied when priorities or comparisons among different countries are investigated. While further research is required, it will be interesting to investigate whether emphasis on different scientific disciplines/domains affects a country's capability to invent and innovate.

South Africa has promoted particularly the advancement of humanities and social sciences fields through the established of the National Institute for the Humanities and Social Sciences (NIHSS) in 2013. The Institute coordinates scholarship, research and ethical practice in these fields within and through the country's public universities. While the impact of the Institute cannot yet be estimated and quantified, the research performance in the fields is an interesting indication that the specific disciplines have grown. Policymakers may have to investigate such opportunities for the technology field too.

References

1. Inglesi-Lotz R, Hakimi A, Pouris A. Patents vs publications and R&D: Three sides of the same coin? Panel Smooth Transition Regression (PSTR) for OECD and BRICS countries. *Appl Econ*. 2018;50(45):4912–4923. <https://doi.org/10.1080/00036846.2018.1468556>
2. Clarivate Analytics. Web of Science database [database on the Internet]. No date [cited 2018 Aug 15]. Available from: <https://clarivate.com/products/web-of-science/databases/>
3. National Advisory Council on Innovation (NACI). South African science, technology and innovation indicators report. Pretoria: NACI; 2017. Available from: http://www.naci.org.za/wp-content/uploads/2018/07/South_African_Science_Technology_And_Innovation_Indicators_Report_2017.pdf





Emergence of vancomycin-resistant enterococci in South Africa: Implications for public health

AUTHORS:

Frank E. Tatsing Foka¹

Ajay Kumar¹

Collins N. Ateba¹

AFFILIATION:

¹Department of Microbiology, School of Biological Sciences, Faculty of Natural and Agricultural Sciences, North-West University, Mafikeng, South Africa

CORRESPONDENCE TO:

Frank Tatsing Foka

EMAIL:

frankeric2023@yahoo.fr

DATES:

Received: 05 Oct. 2017

Revised: 06 Dec. 2017

Accepted: 07 May 2018

Published: 11 Sep. 2018

KEYWORDS:

antimicrobial resistance;
glycopeptide resistance;
emerging pathogens

HOW TO CITE:

Tatsing Foka FE, Kumar A, Ateba CN. Emergence of vancomycin-resistant enterococci in South Africa: Implications for public health. *S Afr J Sci.* 2018;114(9/10), Art. #4508, 7 pages. <https://doi.org/10.17159/sajs.2018/4508>

ARTICLE INCLUDES:

- × Supplementary material
- × Data set

FUNDING:

North-West University

South Africa is among the countries with the highest prevalence of debilitating diseases such as HIV/Aids and diabetes. In this context, the emergence of vancomycin-resistant enterococci (VREs) in most South African ecological niches is quite disturbing, taking into consideration the fact that therapeutic options in a case of resistant-enterococci infection would be limited. Agricultural practices coupled with the misuse of antibiotics in intensive animal rearing and in hospital facilities have led to the creation of reservoirs of VREs in the environment. VREs can cause serious health problems by transmitting their resistance genes to susceptible pathogens; they are transmitted to humans by direct or indirect contact and through the food chain. We screened thoroughly the AJOL and the PubMed databases for studies on VRE incidence in South Africa. This review gives insight into the current status of antimicrobial resistance management in South Africa; it explores the different pathways involved in the spread of VREs and proposes possible solutions to tackle the issue of VREs and antimicrobial resistance in South Africa and other parts of the world.

Significance:

- The recent detection of vancomycin-resistant enterococci in most South African ecological niches poses a serious threat to public health and is therefore an issue of great concern.
- This study not only addresses the causes and patterns of resistance to antimicrobial agents, particularly in South Africa, but also outlines a holistic approach to potential strategies to tackle antimicrobial resistance in South Africa and the world at large.

Introduction

Shortly after antibiotics were introduced for therapeutic purposes, their growth-promoting attributes were discovered; ever since, most antibiotics and their analogues have been used as growth promoters in animal farming.¹ Growth promoters are believed to improve feed conversion, promote animal growth and reduce mortality and morbidity rates resulting from clinical and subclinical illnesses, although the mechanisms through which these effects are achieved are still poorly understood.² This effect of antibiotics motivated the use of avoparcin as a growth promoter for many decades before it was banned worldwide as a result of the emergence of vancomycin-resistant enterococci (VREs).³ The rise of VREs reduced the efficacy of enterococcal infection treatments with teicoplanin and vancomycin (which were the drugs of choice until then), making treatment more challenging.

The isolation of VREs in hospitals and environmental samples worldwide and specifically in South Africa is a serious health concern.⁴⁻⁶ In fact, VREs were isolated from surface water in the Netherlands, from ground water and from hospital waste water in South Africa.^{4,7} VREs mainly cause illnesses in immunocompromised hosts and in patients who are admitted to intensive care units for lengthy periods.⁸ For instance, VREs are responsible for endocarditis, urinary tract infections, bacteraemia, intra-abdominal and pelvic infections, and burn wound and deep tissue infections.

Because of the serious implications of VREs on public health, the issue of VREs cannot be underestimated. This review is therefore not just a citation of reports of VREs in South Africa and their patterns of spread and dissemination, but also an insight into the current management of antimicrobial resistance issues in South Africa with some recommendations on how to tackle the issue of vancomycin resistance genes in South Africa and other countries.

Mechanism of resistance to vancomycin

Two types of vancomycin resistance in enterococci have been reported so far: intrinsic and acquired resistance. Intrinsic resistance refers to an antimicrobial drug being ineffective as a result of inherent features in a species, like restricting drug accessibility to the target or not having the drug target. Acquired resistance occurs when the bacterium is initially susceptible, but develops resistance either by somatic mutation or by acquisition of genes by horizontal transfer. Characterised by low-level resistance to vancomycin, intrinsic resistance is commonly detected in *Enterococcus gallinarum*, *E. casseliflavus* and *E. flavescens*. As opposed to *E. faecium* and *E. faecalis* and less often *E. raffinosus*, *E. avium* and *E. durans* display acquired resistance to vancomycin resulting from the acquisition of genetic determinants either from another organism or from the environment.⁹⁻¹¹ Resistance to vancomycin is conferred by 10 gene clusters: *vanA*, *vanB*, *vanC*, *vanD*, *vanE*, *vanF*, *vanG*, *vanL*, *vanM* and *vanN*. Gene clusters *vanA*, *vanB*, *vanC*, *vanD*, *vanE* and *vanG* have been extensively studied (Figure 1). The resistance phenotypes share the same basic mechanism of resistance. The glycopeptides bind to the carboxy-terminal D-Ala residues of cell wall precursors, thus preventing their incorporation into the nascent peptidoglycan. Substituting the terminal D-Ala residue with either D-lactate (*vanA*, *vanB* and *vanD* genotypes) or D-serine (*vanC* and *vanE* genotypes) confers resistance (Figure 1).^{11,12}

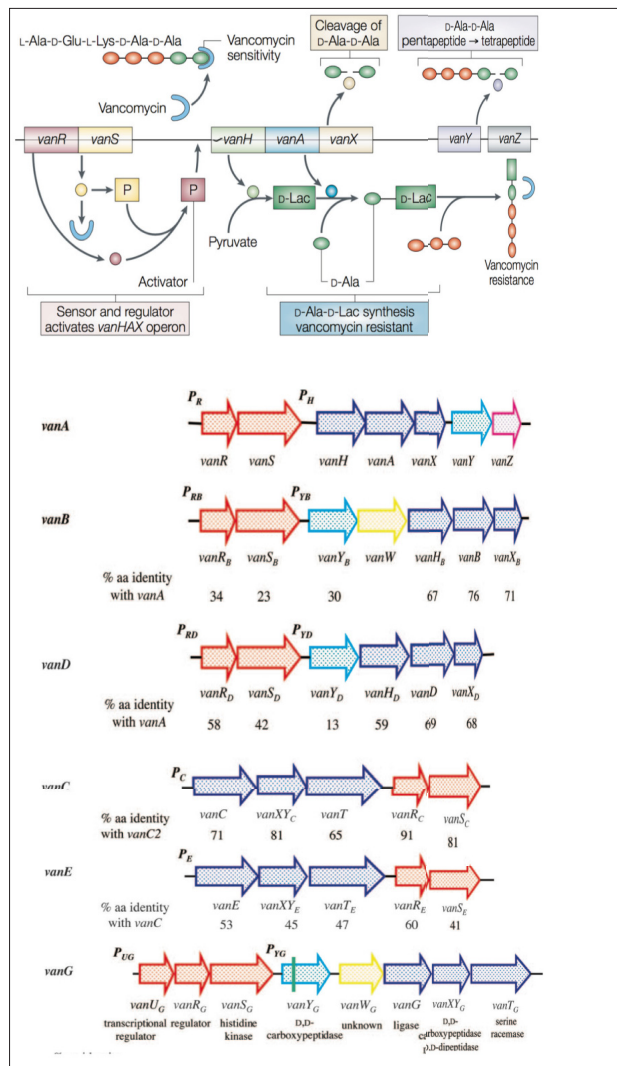


Figure 1: Vancomycin resistance gene clusters and resistance mechanism.¹³

Studies of vancomycin-resistant enterococci in South Africa

Vancomycin-resistant enterococci in food items and the environment

Very few reviews actually give an account of the situation in South Africa with respect to VREs in food items and the environment. Enterococci were isolated from lettuce and spinach leaves in the North West Province using polymerase chain reaction techniques.¹⁴ That study was motivated by the fact that enterococci had previously been screened from ground water intended for drinking in the same area.^{5,6} Vegetables were contaminated as a result of their proximity to the soil when they were grown.¹⁵ Animals and birds were another possible source of contamination but the water used in the irrigation process of these vegetables was a source of contamination. Moreover, the harvesting, packaging, handling and retail hygiene practices were other sources of contamination.¹⁵ As similar findings were made in Oman¹⁶, detection of VREs in fresh, leafy food items constitutes a serious health issue because lettuce is mostly eaten raw, in the form of salads.

Vancomycin-resistant enterococci were isolated from ground water intended for drinking in rural communities of the North West Province.⁵ The investigations indicated that limited access to proper sanitary facilities and lack of hygiene were the causative factors because the water sources were contaminated by faecal matter. These factors are the same as those that contributed to the contamination by VREs of municipal tap water in Mafikeng households, also in the North West Province.¹⁷ In addition to these causes, contamination of environmental

soil and water bodies by resistant isolates can arise as a result of farming activities and contamination from other sources such as contaminated waste effluents from clinical settings.^{4,18,19}

Vancomycin-resistant enterococci in hospital settings

Vancomycin-resistant enterococci constitute a serious issue in South African hospitals, especially taking into consideration the high incidence of HIV-positive patients and the high prevalence of other debilitating illnesses such as tuberculosis and diabetes. A possible synergistic association between Aids or any of the above-mentioned illnesses and VRE infections is rather unsettling and dreaded. In fact, clinical isolates of VREs were reported in South Africa for the first time in 1997.²⁰ A few years later, a case report of VRE infections was filed at the paediatric oncology ward of a tertiary-level paediatric hospital in Cape Town and in the haematology unit of a similar type of hospital in Durban.¹⁹⁻²² Based on the results obtained from gene-sequencing assays (PFGE and MLST), it was suggested that a possible transfer of VRE isolates between patients or persistence of this isolate within the haematology/oncology unit were responsible.¹⁹ This resistance resulted from previous treatments with broad-spectrum antibiotics (mostly third-generation cephalosporins and carbapenems), prolonged stays in the hospital setting, vancomycin exposure, immunosuppression therapy and young age in the case of children.¹⁹⁻²² Moreover, in a recent study, the presence of VREs was reported in wastewater effluents from a hospital facility in Alice, Eastern Cape Province.⁴ These findings highlight the poor or inefficient water-treatment system of that facility's effluents before discharge into the environment. Antibiotics do not undergo total biodegradation in wastewater management processes; this hospital waste water was therefore regarded as a reservoir of resistant pathogens and an ideal environment for exchange of resistance genes or transfer to non-resistant isolates.²³⁻²⁵ These findings therefore demonstrate that the Victoria hospital waste water was a significant source of VREs in the wastewater treatment plant of Fort Hare. As VREs were screened from cow dung and environmental water sources in three selected dairy farms in the Amathole District, waste water from Victoria hospital must have a huge impact on the microbiological quality of the neighbouring environmental water bodies.²⁶

Vancomycin-resistant enterococci in farming and agricultural practices

The Eastern Cape Province and part of the North West Province are mostly agrarian areas with countless numbers of piggery, poultry and cattle farms. Antimicrobials are used for the enhancement of productivity in South Africa.^{27,28} The shedding of resistant bacteria in the environment through faecal contamination is a concern. VREs that possessed *vanB* and *vanC1/C2/C3* resistance genes were screened in pig dung in the Eastern Cape.²⁸ Antibiotic resistance genes were spread in the environment as a result of excretion, flushing of out-of-date prescriptions, medical waste, leakage of septic tanks, effluents from wastewater treatment plants and agricultural waste discharges.²⁸ The subtherapeutic doses of antimicrobials added in the animal feeds for prophylactic purposes and growth-promoting effects were the major causes of the resistance observed in the isolates.^{29,30} In fact, growth promoters have been associated worldwide with the rise of resistant bacteria, leading to the banning of most growth promoters.^{1-3,27}

Pathways of antimicrobial resistance transmission

Antimicrobial agents are not fully transformed into inactive compounds in the systems of treated animals and are excreted in manure where they revert to their initial state after some time.³¹ This makes manure a hotspot for isolates carrying mobile genetic resistance elements and when mixed with soil for agricultural purposes, antibiotic resistance genes are likely to be vertically and horizontally transferred to soil bacteria. This transference leads to the pollution of soil by antibiotic resistance genes and aids in the uptake of antimicrobial resistance by commensal bacteria and human pathogens such as enterococci.³²⁻³⁴ Moreover, treated waste water is known to harbour antibiotic resistance genes which can be transferred to vegetables and crops through irrigation water³⁵ (Figure 2).

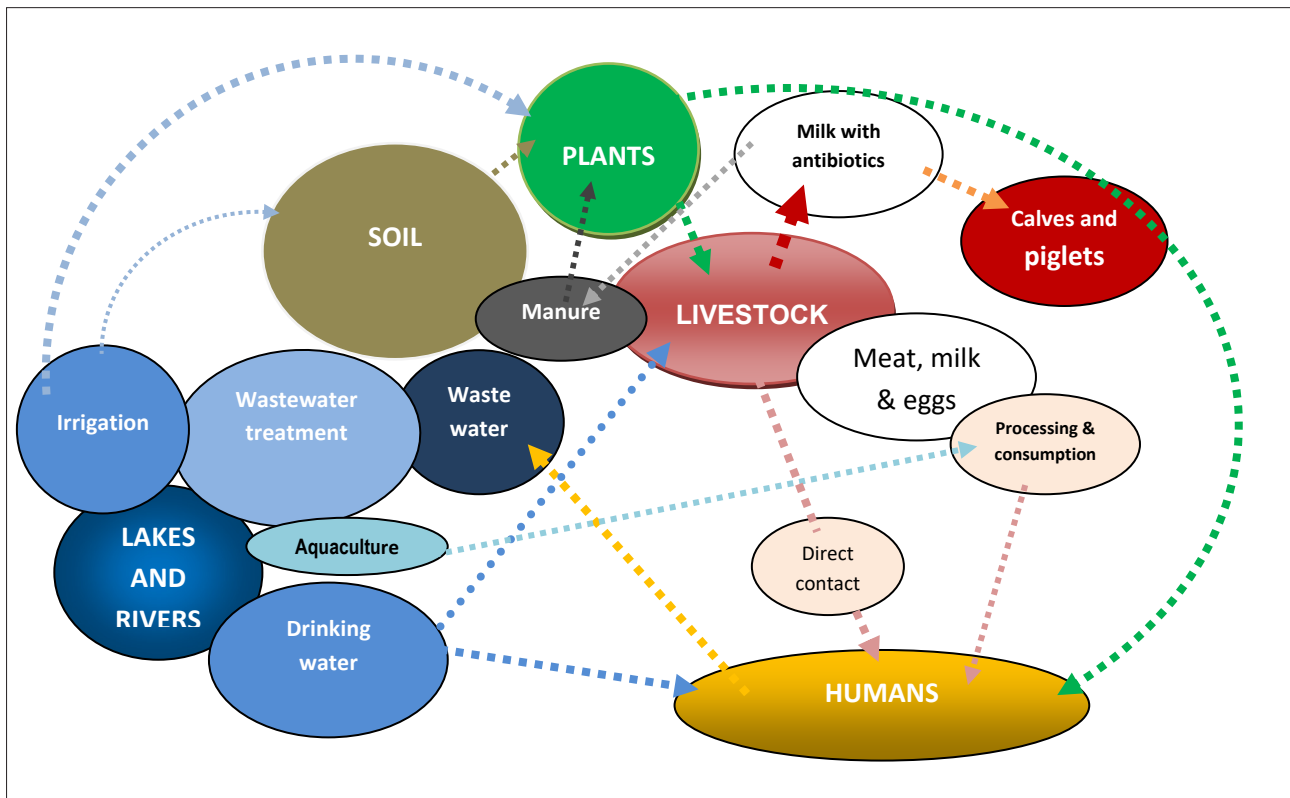


Figure 2: Transmission pathway of antimicrobial resistance within agriculture, the environment and the food-processing industry.

There are several means through which resistant bacteria spread to humans, whether it be from food or from farm animals (Figure 2). The most plausible ones are via the food chain; direct or indirect contact with persons working in close proximity with animals such as farmers and animal healthcare professionals; and environmental components contaminated by agricultural waste or manure.^{2,32-38} Environment plays a key role as a potential reservoir of resistance genes whereby genetic determinants are exchanged amongst isolates that are taken up by humans and animals.^{4,39} Also, prolonged stay in a hospital setting can lead to patients acquiring VREs.⁴

Minimally processed food items, and raw and fermented food items constitute a potential risk through considerable numbers of viable cells; these food items could interact with other factors, such as cohabitation with pathogens, leading to the appearance of resistant strains of enterococci in the human gastrointestinal tract. The *in-vitro* transfer of erythromycin resistance genes from lactic acid bacteria to *Listeria monocytogenes* has been demonstrated.⁴⁰⁻⁴² Moreover, numerous reports have highlighted the transfer of resistance genes amongst isolates of the same species in the human gut.⁴³

Current status of the management of antimicrobial resistance in South Africa

Although there have been reports of the occurrence of antimicrobial-resistant genes in South African livestock and food items, the quantities of antimicrobials used are not yet monitored. Despite being banned a long time ago, some antimicrobial agents are still used in animal rearing (Table 1). Providing data on the consumption of antimicrobials is vital in the assessment and/or management of antimicrobial resistance, but pharmaceutical companies keep secret their data on the amount of antimicrobials sold in the South African market, not to mention the numerous varieties that are available to farmers over the counter.

Table 1: Antimicrobials used in South Africa as growth promoters

Antibiotic	Banned since	Antibiotic group	Related therapeutics	Mechanism of action
Bambermycin		Glycolipid		Inhibition of cell wall synthesis
Bacitracin	1999	Cyclic peptide	Bacitracin	Inhibition of cell wall synthesis
Monensin		Ionophore		Disintegration of cell membrane
Salinomycin		Ionophore		Disintegration of cell membrane
Virginiamycin	1999	Streptogramin	Quinupristin/dalfopristin	Inhibition of protein synthesis
Tylosin	1999	Macrolide	Erythromycin and others	Inhibition of protein synthesis
Spiramycin	1999	Macrolide	Erythromycin and others	Inhibition of protein synthesis
Avoparcin	1997	Glycopeptide	Vancomycin/teicoplanin	Inhibition of cell wall synthesis
Avilamycin		Orthosomycin	Evernimomycin	Inhibition of protein synthesis
Ardacin	1997	Glycopeptide	Vancomycin/teicoplanin	Inhibition of cell wall synthesis

Source: Moyane et al. ²⁷

South Africa is part of the Global Resistance Partnership (GARP) launched in February 2010. The intention of this Partnership is to address and analyse antimicrobial resistance issues in South Africa and partnering countries. Moreover, the South African National Veterinary Surveillance and Monitoring Programme for Resistance to Antimicrobial Drugs (SANVAD) was created alongside the South African Antibiotic Stewardship Programme (SAASP). Reports produced by these entities are still yet to be transcribed into fully operational policies and action plans. However, South Africa remains the most active African country as far as antimicrobial resistance surveillance is concerned. In this regard, the Global Action Plan on antimicrobial resistance initiated in partnership with the World Health Organization emphasises optimisation coupled with a strengthening of the knowledge and evidence base through surveillance and research on antimicrobial usage in human and animals. However, there still exists a paucity of consumption data worldwide, including in South Africa.⁴⁴ Reports of the situational analysis of antibiotic use and resistance carried out in 2011 by South African stakeholders and the GARP through the Centre for Disease Dynamics Economics and Policy (CDDEP) project revealed many setbacks in the antimicrobial stewardship programmes; amongst which there was the unavailability of data from the intercontinental marketing services, causing a bias in the real picture of antimicrobial consumption in South Africa.

Regulation of antimicrobial usage in humans

Prescriptions in the public sector are guided by the Standard Treatment Guidelines which are based on the inclusion and the availability of the medicines on the Essential Medicines List.⁴⁴ Prescriptions are unrestricted in the private sector with prescribers selecting whatever

antimicrobials they feel are most appropriate. Provision is made through the *Nursing Act 33 of 2005* under Section 56(6) in such a way that nurses can also prescribe to patients, especially in public HIV and TB healthcare centres. *Act No. 53 of 1974* allows pharmacists to diagnose and prescribe antibiotics to patients if they suffer from common illnesses.⁴⁵ A comparison of public and private sector data from the past 3 years from Intercontinental Marketing Services reveals an increase in the consumption of certain antibiotics (Table 2). This increase was a result of inappropriate use of antibiotics as first-line medicines and a lack of awareness of appropriate antibiotic prescribing or simply a willingness to use new drug formulations based on their availability on the market or based on the emergence of resistance to previous antibiotics.

Regulation of antimicrobial usage in animals

The Department of Agriculture, Forestry and Fisheries and the National Department of Health regulate the use of antibiotics by administering the *Fertilizers, Farm Feeds, Agricultural Remedies and Stock Remedies Act (Act 36 of 1947)* and the *Medicines and Related Substances Control Act (Act 101 of 1965)*, respectively. The *Stock Remedies Act 36 of 1947* was initiated to control the numerous parasites that were infesting South African livestock at the time. As time went by, certain antimicrobials, such as growth promoters, were allowed in order to help farmers in rural areas to access essential livestock medicines. The *Medicines and Related Substances Control Act 101 of 1965* was initiated for prescription-only medicines and veterinary antibiotics are also controlled by this Act.

Table 2: Antibiotics sold in the private and public sector from 2014 to 2016

Antibiotic class	MAT units ^a (2014)		MAT units ^a (2015)		MAT units ^a (2016)		CAGR ^b (2014–2016)		% Market share (2016)	
	Private	Public	Private	Public	Private	Public	Private	Public	Private	Public
Tetracycline + combinations	307 170	226 993 000	296 428	114 988 400	282 220	168 296 842	-3%	21%	1%	9%
Chloramphenicol + combinations	1124	121	983	109	1014	93	-3%	-8%	0%	0%
Broad-spectrum penicillin oral	8 249 655	530 513 290	8 607 223	54 045 080	7 826 870	385 061 012	-2%	167%	35%	20%
Broad-spectrum penicillin injectable	520 470	49 241 030	533 780	34 687 670	572 498	54 987 307	3%	26%	3%	3%
Cephalosporin oral	1 951 706	12 221 600	1 854 653	1 455 300	1 705 486	7 300 010	-4%	124%	8%	0%
Cephalosporin injectable	2 053 062	16 097 300	2 036 180	10 565 000	2 015 283	76 629 057	-1%	169%	9%	4%
Trimethoprim combinations	1 466 062	966 535	1 491 648	783 509 493	1 437 019	700 365 086	-1%	-5%	6%	37%
Macrolides and similar types	2 822 661	185 162 200	2 935 812	8 019 700	2 874 181	16 427 840	1%	43%	13%	1%
Oral fluoroquinolone	3 618 738	11 995 000	3 576 474	23 465 600	3 378 464	33 679 945	-2%	20%	15%	2%
Injectable fluoroquinolone	641 067	58 481	659 363	144 100	560 007	2 158 000	-4%	287%	3%	0%
Aminoglycosides	79 908	6578 500	87 101	6 975 300	89 754	6 295 783	4%	-5%	0%	0%
Penems and carbapenems	1 141 501	1 991 900	1 276 979	460 000	1 093 413	809 878	-1%	33%	5%	0%
Glycopeptides	179 134	257 500	182 071	285 700	190 314	651 093	2%	51%	1%	0%
Medium-/narrow-spectrum penicillins	280 172	626 304 600	292 893	515 183 440	145 960	424 833 433	-20%	-9%	1%	22%
All other antibacterials	21 378	2 530 000	25 991	5 899	33 037	28 704 650	16%	6876%	0%	2%
Grand total	23 333 808	1 670 911 057	23 857 579	1 553 790 791	22 205 520	1 906 200 029	-2%	11%	100%	100%

Source: Schellack et al.⁴⁵

^aMAT, moving annual total, i.e. the total value of the sales figures for the product, over the course of the period displayed.

^bCAGR, compound annual growth rate; this indicator was used as a measure of the market growth over multiple time periods for the two sectors.

Measures to reduce the incidence and prevalence of vancomycin-resistant enterococci

Step 1: Enforcement of the legislation on drug distribution and usage

The dual registration process of medicines through the *Agricultural Remedies and Stock Remedies Act (Act 36 of 1947)* and the *Medicines and Related Substances Control Act (Act 101 of 1965)* presents some flaws and has raised concerns about the exacerbation of antimicrobial resistance emergence if there is no effective control. Stock remedies are distributed to veterinary wholesalers, distributors, farmers' cooperatives and feed mix companies by the manufacturer. Consequently, stock remedies are freely available and no record is kept of their use. Moreover, the South African situation deviates from the 1998 World Health Organization best practice guidelines in that: (1) the dual system of regulating veterinary products only partially addresses clear, transparent manufacturing requirements (whereas antibiotics listed under *Act 101 of 1965* must be authorised with a Good Manufacturing Practice licence, stock remedies under *Act 36 of 1947* are not); and (2) most authorised veterinary antibiotics are over-the-counter stock remedies and often administered by farmers. The World Health Organization recommends that only trained and licensed professionals decide when and how to use antibiotics.

Step 2: Prioritisation of the use of alternatives to antibiotics

Although vaccination campaigns are often costly, they could reduce the amounts of antimicrobials used in farming.^{46,47} Moreover, the use of probiotics, prebiotics and synbiotics in animal farming should be encouraged because these will improve the gut bacterial flora and therefore reduce disease occurrence that could necessitate antimicrobial usage.^{47,48} It has been shown that healthier gut microbial flora contribute to a better immune system with a better nutrient uptake and less colonisation by pathogens.⁴⁹ Studies have demonstrated the usage of predatory bacteria to counter the pathogenic effects of the same strains or other pathogens after oral administration in chickens, cows and rabbits.^{50,51} Moreover, specific yeast strains with specific properties could be used as probiotics.^{52,53} Because of their specificity and their selectivity towards bacterial strains, bacteriophages represent another potential alternative to antimicrobial management of illnesses in animal farming even though they can be used in combination with some antimicrobials as there is no negative interaction between the two treatments.⁵³⁻⁵⁶ In addition, there is a tremendous amount of research on the potential use of antimicrobial peptides such as bacteriocins as an alternative to antibiotics.^{46,57} The South African research stakeholders and the pharmaceutical companies should therefore harness substantial research funding in these novel areas as they represent huge potential for the future.

Step 3: Implementation of a nationwide effective antimicrobial resistance surveillance system

Large databases on antimicrobial resistance can be created and used for risk analyses and management in the different sectors of activity as well as in the different ecological niches that play a role in the dissemination and spread of resistant enterococci. Such surveillance systems have already been put in place in the European Union and in the USA to trace and control the patterns of antimicrobial distribution and that of the dissemination of antimicrobial-resistant bacteria. The database produced is playing a significant impact in the decision-making spheres in order to tackle antimicrobial resistance issues in the EU. This particular measure would be efficient if scientists and researchers were at the centre of the process and if international collaboration was encouraged because of the global aspect of antimicrobial resistance issues. In this regard, 70 experts from 33 countries gathered in 2011 at the Third World Healthcare Associated Infections Forum (WHAIF) that was dedicated to antibiotic resistance awareness and action.⁵⁸ Actions that were agreed at this Forum are summarised in Table 3.

Last but not the least, the adoption of healthy lifestyles and the development of new drugs and vaccines should be encouraged as these constitute new perspectives as far as the therapeutic management of VRE infections is concerned.

Table 3: Actions prioritised at the Third World Healthcare Associated Infections Forum with respect to antimicrobial resistance

Stakeholder	Action to prioritise
Policymakers and health authorities	<ul style="list-style-type: none"> In animals, analogues to human medicines should not be utilised and antibiotics should be utilised only as therapeutics with certain categories of antimicrobials used solely for human therapy
	<ul style="list-style-type: none"> Growth promoters in animal feed should be banned worldwide
	<ul style="list-style-type: none"> Commercialisation of antimicrobials used as human medicines should be regulated and over-the-counter trading should be prohibited in all countries
	<ul style="list-style-type: none"> International organisations (World Health Organization, European Union) should come up with a charter of conduct to which all countries worldwide will abide, as far as the availability of antibiotics is concerned
Human and veterinary healthcare communities	<ul style="list-style-type: none"> Implement surveillance strategies of antimicrobial usage and monitoring systems of emergence and spread of bacterial resistance
	<ul style="list-style-type: none"> Promote specific courses in medical and veterinary schools for training on antimicrobial resistance mechanisms and wise usage of antibiotics, taking into consideration the cultural background of each country
General public	<ul style="list-style-type: none"> Sensitise the general public to the importance of antibiotic protection and wise usage of antimicrobials, that is, only when it is necessary
	<ul style="list-style-type: none"> Improvement of sanitation systems and teaching of basic hygiene practices such as hand washing in order to reduce the spread of illnesses and resistant isolates in the environment
	<ul style="list-style-type: none"> Promote the participation of consumers in the conception and application of the strategies developed
Industrial companies	<ul style="list-style-type: none"> Design reliable and rapid diagnostic tests that could be utilised by a patient or a doctor to guide the prescription of antibiotics and avoid their prescription for viral infections
	<ul style="list-style-type: none"> Promote research and development of new antimicrobials
	<ul style="list-style-type: none"> Determine new economic paradigms that take into consideration public health interests alongside industrial quests for profits

Source: Jarlier et al.⁵⁸

Conclusion

Antimicrobials have been used in human and veterinary medicine ever since they were discovered. The intensive usage of antimicrobials in South Africa with their negative impact on food safety and human health has led to the emergence and the spread of VREs and other antimicrobial-resistant bacteria. In fact, epidemiological studies have demonstrated the link between antimicrobial usage and the detection of resistant isolates in the environment, making therapeutic alternatives for the treatment of infections caused by such organisms limited and their subsequent antimicrobial resistance genes available in the environment. As VREs are becoming a serious health threat in South Africa and worldwide, there is

an urgent need to address this issue. Although the use of alternatives to antibiotics will effectively slow down the emergence process of resistant bacteria such as VREs, a synergistic approach involving the revision and the enforcement of laws and regulations on drug usage and distribution in South Africa, the encouragement of the invention of novel therapeutic molecules and the implementation of a critical nationwide antimicrobial resistance surveillance system will have significant impacts on the reduction of the prevalence or incidence of VREs as well as that of other antimicrobial-resistant bacteria. As awareness among the scientific community and the major stakeholders is increasing in South Africa, combined international action is required to tackle this issue once and for all.

Limitations

A limitation of this review was the lack of animal antimicrobial consumption data because of the unavailability or scarcity of data from wholesale suppliers and the Department of Agriculture, Forestry and Fisheries of South Africa.

Acknowledgements

We acknowledge the financial support of the North-West University towards the realisation of this study.

Authors' contributions

F.E.T.F. carried out the study and wrote the manuscript; A.K. and C.N.A. supervised the study and proofread the manuscript.

References

- Butaye P, Devriese LA, Haesebrouck F. Antimicrobial growth promoters used in animal feed: Effects of less well known antibiotics on gram positive bacteria. *Clin Microbiol Rev.* 2003;16(2):175–188. <https://doi.org/10.1128/CMR.16.2.175-188.2003>
- Marshall BM, Levy SB. Food animals and antimicrobials: Impacts on human health. *Clin Microbiol Rev.* 2011;24(4):718–733. <https://doi.org/10.1128/CMR.00002-11>
- Rosvoll TCS, Pedersen T, Sletvold H, Johnsen PJ, Sollid JE, Simonsen GS, et al. PCR-based plasmid typing in *Enterococcus faecium* strains reveals widely distributed pRE25-, pRUM-, pIP501- and pHT β -related replicons associated with glycopeptide resistance and stabilizing toxin-antitoxin systems. *FEMS Immunol Med Microbiol.* 2010;58:254–268. <https://doi.org/10.1111/j.1574-695X.2009.00633.x>
- Iweriebor BC, Gaqavu S, Obi CL, Nwodo UU, Okoh AI. Antibiotic susceptibilities of enterococcus species isolated from hospital and domestic wastewater effluents in Alice, Eastern Cape Province of South Africa. *Int J Environ Res Publ Health.* 2015;12:4231–4236. <https://doi.org/10.3390/ijerph120404231>
- Ateba CN, Maribeng MD. Detection of *Enterococcus* species in groundwater from some rural communities in the Mmabatho area, South Africa: A risk analysis. *Afr J Microbiol Res.* 2011;5(23):3930–3935. <https://doi.org/10.5897/AJMR11.645>
- Ateba CN, Lekoma KP, Kawadza DT. Detection of vanA and vanB genes in vancomycin-resistant enterococci (VRE) from groundwater using multiplex PCR analysis. *J Water Health.* 2013;11(4):684–691. <https://doi.org/10.2166/wh.2013.037>
- Taucer-Kapteijin M, Hoogenboezem W, Heiligers L, Danny de Bolster H, Medema G. Screening municipal wastewater effluent and surface water used for drinking water production for the presence of ampicillin and vancomycin resistant enterococci. *Int J Hyg Environ Health.* 2016;16:1–7.
- Kuzucu C, Cizmeci Z, Durmaz R, Durmaz E, Ozerol IH. The prevalence of faecal colonization of enterococci, the resistance of the isolates to ampicillin, vancomycin and high-level aminoglycosides and the clonal relationship among isolates. *Microb Drug Resist Mech Epidemiol Dis.* 2005;11(2):159–164. <https://doi.org/10.1089/mdr.2005.11.159>
- Sónia R, Ingrid C, Nuno S, Michel H, Hugo S, José-Luis CM, et al. Effect of vancomycin on the proteome of the multiresistant *Enterococcus faecium* SU18 strain. *J Proteom.* 2015;113:378–387. <https://doi.org/10.1016/j.jprot.2014.10.012>

- Jui-Chang T, Po-Ren H, Hsiao-Mann L, Hui-Jen C, Shen-Wu H, Lee-Jene T. Identification of clinically relevant *Enterococcus* species by direct sequencing of groES and spacer region. *J Clin Microbiol.* 2005;43:235–241. <https://doi.org/10.1128/JCM.43.1.235-241.2005>
- Arthur M, Reynolds P, Courvalin P. Glycopeptide resistance in enterococci. *Trends Microbiol.* 1996;4:401–407. [https://doi.org/10.1016/0966-842X\(96\)10063-9](https://doi.org/10.1016/0966-842X(96)10063-9)
- Bugg TD, Wright GD, Dutka-Malen S, Arthur M, Courvalin P, Walsh CT. Molecular basis for vancomycin resistance in *Enterococcus faecium* BM4147: Biosynthesis of a depsipeptide peptidoglycan precursor by vancomycin resistance proteins VanH and VanA. *Biochemistry.* 1991;30:10408–10415. <https://doi.org/10.1021/bi00107a007>
- Hughes D. Exploiting genomics, genetics and chemistry to combat antibiotic resistance. *Nat Rev Genet.* 2003;4:432–441. <https://doi.org/10.1038/nrg1084>
- Ateba CN, Mohapi MI. Isolation of vancomycin resistant enterococci isolated from leafy vegetables (lettuce) from North West Province. *Life Sci.* 2013;10(4):1163–1170.
- Ruimy R, Brisabois A, Bernede C. Organic and conventional fruits and vegetables contain equivalent counts of Gram-negative bacteria expressing resistance to antibacterial agents. *Environ Microbiol.* 2010;12(3):608–615. <https://doi.org/10.1111/j.1462-2920.2009.02100.x>
- Al-Kharousi ZS, Guizani N, Al-Sadi AM, Al-Bulushi IM, Shaharoon B. Hiding in fresh fruits and vegetables: Opportunistic pathogens may cross geographical barriers. *Int J Microbiol.* 2016; Art. #4292417, 14 pages. <https://doi.org/10.1155/2016/4292417>
- Matlou PD. Molecular analysis of vancomycin-resistant enterococci isolated from ground and surface water [MSc thesis]. Mafikeng: North-West University; 2016.
- Molale LG. Characterization of enterococci isolates from water sources in the North West Province [PhD thesis]. Potchefstroom: North-West University; 2016.
- Lochan H, Moodley C, Bamford C, Hendricks M, Davidson A, Eley B. Emergence of vancomycin-resistant *Enterococcus* at a tertiary paediatric hospital in South Africa. *S Afr Med J.* 2016;106(6):562–566. <https://doi.org/10.7196/SAMJ.2016.v106i6.10858>
- Mahabeer Y, Lowman W, Govind CN, Swe-Swe-Han K, Mlisana KP. First outbreak of vancomycin-resistant *Enterococcus* in a haematology unit in Durban, South Africa. *S Afr J Infect Dis.* 2016;31(1):20–24. <https://doi.org/10.1080/23120053.2015.1118819>
- Budavari SM, Saunders GL, Liebowitz LD, Khoosal M, Crewe-Brown HH. Emergence of vancomycin resistant enterococci in South Africa. *S Afr Med J.* 1997;87(11):1557.
- Singh J, Esparza S, Patterson M, Vogel K, Patel B, Gornick W. Vancomycin-resistant *Enterococcus* in paediatric oncology patients: Balancing prevention and family centered care. *J Pediatr Hematol Oncol.* 2013;35(3):227–231. <https://doi.org/10.1097/MPH.0b013e318257a6ca>
- Ergani-Ozcan A, Naas T, Baysan BO, Ogunc D, Inan D, Colak D, et al. Nosocomial outbreak of vancomycin-resistant *Enterococcus faecium* in a paediatric unit at a Turkish university hospital. *J Antimicrob Chemother.* 2008;61(5):1033–1039. <https://doi.org/10.1093/jac/dkn066>
- Isogai N, Urushibara N, Kawaguchiya M, Ghosh S, Suzuki K, Watanabe N, et al. Characterization of *Enterococcus faecium* with macrolide resistance and reduced susceptibility to quinupristin/dalfopristin in a Japanese hospital: Detection of extensive diversity in erm(B)-regulator regions. *Microb Drug Resist.* 2013;19:298–307. <https://doi.org/10.1089/mdr.2012.0176>
- Moges F, Endris M, Belyhun Y, Worku W. Isolation and characterization of multiple drug resistance bacterial pathogens from waste water in hospital and non-hospital environments, northwest Ethiopia. *BMC Res Notes.* 2014;7:1–6. <https://doi.org/10.1186/1756-0500-7-215>
- Ngu TG. Genotypic and phenotypic characterization of enterococci from cow dung and environmental water sources in three selected dairy farms in Amathole District [MSc thesis]. Alice: University of Fort Hare; 2016.
- Moyane JN, Jideani AIO, Aiyegoro OA. Antibiotic usage in food producing animals in South Africa and impact on human: antibiotic resistance. *Afr J Microbiol Res.* 2013;7(24):2990–2997. <https://doi.org/10.5897/AJMR2013.5631>

28. Iweriebor BC, Obi CL, Okoh AI. Virulence and antimicrobial resistance factors of *Enterococcus* spp. isolated from faecal samples from piggery farms in Eastern Cape, South Africa. *BMC Microbiol.* 2015;15, Art. #136, 11 pages. <https://doi.org/10.1186/s12866-015-0468-7>
29. Jackson CR, Fedorka-Cray PJ, Barrett JB, Ladely SR. High-level aminoglycoside resistant enterococci isolated from swine. *Epidemiol Infect.* 2005;133:367–371. <https://doi.org/10.1017/S0950268804003395>
30. Aarestrup FM, Kruse H, Tast E, Hammerum AM, Jensen LB. Associations between the use of antimicrobial agents for growth promotion and the occurrence of resistance among *Enterococcus faecium* from broilers and pigs in Denmark, Finland, and Norway. *Microb Drug Resist.* 2000;6:63–70. <https://doi.org/10.1089/mdr.2000.6.63><https://doi.org/10.1089/mdr.2000.6.71>
31. Boxall AB, Blackwell P, Cavallo R, Kay P, Tolls J. The sorption and transport of a sulphonamide antibiotic in soil systems. *Toxicol Lett.* 2002;131:19–28. [https://doi.org/10.1016/S0378-4274\(02\)00063-2](https://doi.org/10.1016/S0378-4274(02)00063-2)
32. Ding GC, Radl V, Schloter-Hai B, Jechalke S, Heuer H, Smalla K, et al. Dynamics of soil bacterial communities in response to repeated application of manure containing sulfadiazine. *PLoS One.* 2014;9(3), e92958, 10 pages. <https://doi.org/10.1371/journal.pone.0092958>
33. Forsberg KJ, Patel S, Gibson MK, Lauber CL, Knight R, Fierer N, et al. Bacterial phylogeny structures soil resistomes across habitats. *Nature.* 2014;509:612–616. <https://doi.org/10.1038/nature13377>
34. Thanner S, Drissner D, Walsh F. Antimicrobial resistance in agriculture. *MBio.* 2016;7(2), e02227-15, 7 pages. <https://doi.org/10.1128/mBio.02227-15>
35. Drissner D, Zürcher U. Microbial safety of fresh fruits and vegetables. In: Motarjemi Y, editor. *Encyclopedia of food safety.* Vol 3. Oxford: Elsevier; 2014. <https://doi.org/10.1016/B978-0-12-378612-8.00280-8>
36. Soonthornchaikul N, Garelick H. Antimicrobial resistance of *Campylobacter* species isolated from edible bivalve molluscs purchased from Bangkok markets, Thailand. *Foodborne Pathog Dis.* 2009;6(8):947–951. <https://doi.org/10.1089/fpd.2008.0236>
37. Petersen A, Andersen JS, Kaewmack T, Somsiri T, Dalsgaard A. Impact of integrated fish farming on antimicrobial resistance in a pond environment. *Appl Environ Microbiol.* 2002;68(12):6036–6042. <https://doi.org/10.1128/AEM.68.12.6036-6042.2002>
38. Shah SQA, Colquhoun DJ, Nikuli HL, Sørum H. Prevalence of antibiotic resistance genes in the bacterial flora of integrated fish farming environments of Pakistan and Tanzania. *Environ Sci Technol.* 2012;46(16):8672–8679. <https://doi.org/10.1021/es3018607>
39. Rossi F, Rizzotti L, Felis GE, Torriani S. Horizontal gene transfer among microorganisms in food: Current knowledge and future perspectives. *Food Microbiol.* 2014;42:232–243. <https://doi.org/10.1016/j.fm.2014.04.004>
40. Toomey N, Monaghan A, Fanning S, Bolton DJ. Assessment of antimicrobial resistance transfer between lactic acid bacteria and potential foodborne pathogens using in vitro methods and mating in a food matrix. *Foodborne Pathog Dis.* 2009;6(8):925–933. <https://doi.org/10.1089/fpd.2009.0278>
41. Doucet-Populaire F, Trieu-Cuot P, Dosbaa I, Andreumont A, Courvalin P. Inducible transfer of conjugative transposon Tn1545 from *Enterococcus faecalis* to *Listeria monocytogenes* in the digestive tracts of gnotobiotic mice. *Antimicrob Agents Chemother.* 1991;35(1):185–187. <https://doi.org/10.1128/AAC.35.1.185>
42. Rizzotti L, La Gioia F, Dellaglio F, Torriani S. Molecular diversity and transferability of the tetracycline resistance gene tet(M), carried on Tn916-1545 family transposons, in enterococci from a total food chain. *Antimicrob Agents Chemother.* 2009;96(1):43–52. <https://doi.org/10.1007/s10482-009-9334-7>
43. Economou V, Gousia P. Agriculture and food animals as a source of antimicrobial-resistant bacteria. *Infect Drug Resist.* 2015;8:49–61. <https://doi.org/10.2147/IDR.S55778>
44. Perumal-Pillay V, Suleman F. Selection of essential medicines for South Africa — an analysis of in-depth interviews with national essential medicines list committee members. *BMC Health Serv Res.* 2017;17, Art.#17, 17 pages. <https://doi.org/10.1186/s12913-016-1946-9>
45. Schellack N, Deon B, Brink A, Duse A, Faure K, Goff D, et al. A situational analysis of current antimicrobial governance, regulation, and utilization in South Africa. *Int J Infect Dis.* 2017;64:100–106. <https://doi.org/10.1016/j.ijid.2017.09.002>
46. Allen HK, Trachsel J, Looft T, Casey TA. Finding alternatives to antibiotics. *Ann NY Acad Sci.* 2014;1323(1):91–100. <https://doi.org/10.1111/nyas.12468>
47. Callaway TR, Edrington TS, Anderson RC. Probiotics, prebiotics and competitive exclusion for prophylaxis against bacterial disease. *Anim Health Res Rev.* 2008;9(2):217–225. <https://doi.org/10.1017/S1466252308001540>
48. Gaggia F, Mattarelli P, Biavati B. Probiotics and prebiotics in animal feeding for safe food production. *Int J Food Microbiol.* 2010;141(1):15–28. <https://doi.org/10.1016/j.ijfoodmicro.2010.02.031>
49. Seal BS, Lillehoj HS, Donovan DM, Gay CG. Alternatives to antibiotics: A symposium on the challenges and solutions for animal production. *Anim Health Res Rev.* 2013;14(1):78–87. <https://doi.org/10.1017/S1466252313000030>
50. Atterbury RJ, Hobbly L, Till R. Effects of orally administered *Bdellovibrio bacteriovorus* on the well-being and *Salmonella* colonization of young chicks. *Appl Environ Microbiol.* 2011;77(16):5794–5803. <https://doi.org/10.1128/AEM.00426-11>
51. Dwidar M, Monnappa AK, Mitchell RJ. The dual probiotic and antibiotic nature of *Bdellovibrio bacteriovorus*. *BMB Rep.* 2012;45(2):71–78. <https://doi.org/10.5483/BMBRep.2012.45.2.71>
52. Biliouris K, Babson D, Schmidt-Dannert C, Kaznessis YN. Stochastic simulations of a synthetic bacteria-yeast ecosystem. *BMC Syst Biol.* 2012;6(1), Art. #58, 13 pages. <http://dx.doi.org/10.1186/1752-0509-6-58>
53. Kenny M, Smidt H, Mengheri E, Miller B. Probiotics – do they have a role in the pig industry? *Animal.* 2011;5(3):462–470. <https://doi.org/10.1017/S175173111000193X>
54. Chan BK, Abedon ST, Loc-Carrillo C. Phage cocktails and the future of phage therapy. *Future Microbiol.* 2013;8(6):769–783. <https://doi.org/10.2217/fmb.13.47>
55. Goodridge LD, Bisha B. Phage-based biocontrol strategies to reduce foodborne pathogens in foods. *Bacteriophage.* 2011;1(3):130–137. <https://doi.org/10.4161/bact.1.3.17629>
56. Balogh B, Jones JB, Iriarte FB, Momol MT. Phage therapy for plant disease control. *Curr Pharm Biotechnol.* 2010;11(1):48–57. <https://doi.org/10.2174/138920110790725302>
57. Snyder AB, Worobo RW. Chemical and genetic characterization of bacteriocins: Antimicrobial peptides for food safety. *J Sci Food Agric.* 2014;94(1):28–44. <https://doi.org/10.1002/jsfa.6293>
58. Jarlier V, Carlet J, McGowan J, participants of the 3rd World Healthcare-Associated Infections Forum. Priority actions to fight antibiotic resistance: Results of an international meeting. *Antimicrob Resist Infect Control.* 2012;1(1), Art. #17, 3 pages. <https://doi.org/10.1186/2047-2994-1-17>





Aetiological doctrines and prevalence of pellagra: 18th century to middle 20th century

AUTHORS:

Margaretha Viljoen¹
Priyesh Bipath²
Johannes L. Roos¹

AFFILIATIONS:

¹Department of Psychiatry,
University of Pretoria, Pretoria,
South Africa

²Department of Physiology,
University of Pretoria, Pretoria,
South Africa

Correspondence to:

Margaretha Viljoen

EMAIL:

mviljoen@webafrica.org.za

DATES:

Received: 23 Aug. 2017

Revised: 28 Feb. 2018

Accepted: 11 May 2018

Published: 11 Sep. 2018

KEYWORDS:

maize; niacin; tryptophan;
nutritional-deficiency disease

HOW TO CITE:

Viljoen M, Bipath P, Roos JL.
Aetiological doctrines and
prevalence of pellagra: 18th
century to middle 20th century.
S Afr J Sci. 2018;114(9/10),
Art. #4597, 7 pages. <https://doi.org/10.17159/sajs.2018/4597>

ARTICLE INCLUDES:

- × Supplementary material
- × Data set

FUNDING:

None

Pellagra is characterised by dermatological, gastrointestinal and neuropsychiatric manifestations. Millions contracted the disease and hundreds of thousands died between the time it was first recorded until pellagra was finally recognised as a niacin-deficiency disease. Pellagra became epidemic when maize, with its limited bio-availability of nutrients such as niacin and tryptophan, became the staple food in the near-monophagic diets of the impoverished and institutionalised. By the mid-20th century, pellagra was all but eradicated in large parts. The decline in prevalence can largely be ascribed to a better understanding of the link between nutrition and disease, improvements in socio-economic conditions of workers and food enrichment. We briefly review aetiological doctrines on pellagra and the global spread of the disease from the early 18th century until the middle of the 20th century. In the final analysis, we examine the reasons for, and the legitimacy of, the persistent association between pellagra and the consumption of maize.

Significance:

- Almost two centuries have elapsed since the first description of pellagra and its general acceptance as a nutritional-deficiency disease.
- The link between maize and pellagra is primarily a reflection of the nutritional inadequacies of a near monophagic diet over-dependent on a grain deficient in bioavailable niacin and tryptophan.
- We refute the concept of nixtamalisation as the main reason for the apparent absence of pellagra in early pre-Columbian North American, Mesoamerican and South American cultures.

Introduction

Pellagra is a nutrition-deficiency disease characterised by dermatological, gastrointestinal and neuropsychiatric manifestations. The symptoms are commonly referred to as the three D's: dermatitis, diarrhoea and dementia, which may not necessarily present in that order. Although the symptoms of pellagra are defined as dermatitis, diarrhoea and dementia, a wide range of gastrointestinal and neuropsychiatric symptoms may present.^{1,2} The skin lesions appear primarily on those parts exposed to sunlight, such as the dorsum of the hands and feet, with the typical 'Casal's necklace' of rough skin in exposed areas around the neck and upper chest.¹

Pellagra is the classical syndrome of a severe deficiency of the water-soluble vitamin niacin.² Niacin, also known as nicotinic acid or vitamin B3, and its amide derivative nicotinamide, are precursors of the co-enzyme nicotinamide adenine dinucleotide (NAD) which can be phosphorylated (NADP) and/or reduced (NADH and NADPH).³ As NAD and its derivatives are involved in a multitude of oxidation-reduction as well as non-redox reactions, severe niacin deficiency can lead to death with multi-organ involvement.

Niacin is abundant in a number of different foods – dairy products, cereals, nuts, leafy vegetables, yeast, fish, and meat products.⁴ However, the form in which it occurs and its bio-availability differ. Some foods contain niacin in the free form that is highly bio-available. Cereal grains, as a source of niacin, vary depending on the type and the processes involved in food preparation. In addition to dietary intake, niacin can also be synthesised from the essential amino acid tryptophan via the kynurenine pathway.⁵ Such *de novo* synthesis of niacin decreases with a deficiency in iron, riboflavin (vitamin B2) or pyridoxine (vitamin B6).⁶ Several factors may have pellagrigenic effects, either through their influence on the absorption of niacin and/or tryptophan, or through their effects on the conversion of tryptophan to niacin. These factors include excessive leucine intake⁷, chronic alcoholism^{1,2}, anti-tuberculosis drugs⁸, immunosuppressive drugs⁹, eating disorders and a variety of gastrointestinal disturbances².

Today, the name pellagra has become virtually synonymous with the term niacin deficiency. However, many aetiological theories were explored and almost two centuries elapsed between the first description of the condition and its general acceptance as a nutritional-deficiency disease. We briefly review the main proposed aetiological doctrines, the global spread of the disease from the early 18th century until the middle of the 20th century and the association between pellagra and maize.

Aetiology of pellagra: From vampires to nutritional deficiency

Up until its formal recognition as a dietary deficiency, pellagra has been blamed on various phenomena. It is even believed by some to be the origin of the vampire myth on which Bram Stoker's blood-sucking Transylvanian vampire, Count Dracula, was based. This association was perhaps largely because of the sunlight sensitivity of pellagrins and the belief that vampires were said to avoid sunlight in order to maintain their strength, but also because of shared psychological traits such as insomnia, anxiety, aggression and depression. Similarities between the physical and psychological characteristics of pellagrins and that attributed by folklore to vampires are described in an intriguing overview by Hampl and Hampl¹⁰ based on early 1700s European literature and folklore.

Theories on the aetiology of pellagra can broadly be divided into maize-based theories and those based on factors other than the consumption of maize. The latter includes theories that claimed pellagra was caused by a blood-

sucking insect, the ingestion of semi-dried edible oils and the retention of mineral salts, or was an inherited disease, a race-dependent disorder, a type of scurvy, a disease related to leprosy and, perhaps the view with the most far reaching consequences, a communicable condition caused by infectious agents.¹¹⁻¹⁴ A more comprehensive treatise on these early aetiological doctrines can be found in a 1912 document by Niles¹⁵. Probably most important among the theories that attributed pellagra to factors other than maize is Sambon's infectious paradigm that shaped the early American investigations and hampered investigations and the correct treatment of pellagrins in the USA during the early 1900s.¹⁶ Louis Westenra Sambon (1867–1931) was born in Milan, obtained an MD from the University of Naples and after some award-winning work on cholera moved to England where he spent the major part of his career at the Liverpool School of Tropical Medicine. In line with his groundbreaking theories on sleeping sickness, malaria and yellow fever, Sambon was convinced that pellagra was an infectious disease caused by a protozoan parasite and transmitted by an insect (*Simulium*). This parasitic theory of Sambon's was first introduced in 1905 at the meeting of the British Medical Association.^{12,16}

The most enduring theories about the aetiology of pellagra were based on the consumption of maize, to the extent that an association can still be found in modern medical textbooks.¹ There were several variations of the hypothesis that maize consumption was the cause of pellagra. A theory given prominence by the Italian psychiatrist Cesare Lombroso in 1892, was that pellagra was caused by a toxin in deteriorating maize.^{13,17} Lombroso's toxico-chemical approach, also referred to as the toxic zeist or spoiled maize theory, stated that inadequately dried, stored maize, under the influence of microorganisms, undergoes certain changes with the formation of one or more toxic chemical substances that could lead to pellagra. Lombroso's theory was more explicitly articulated in a 1905 pamphlet by Alpago-Novello, president of the Pellagological Commission of the Province of Belluno and close friend and follower of Lombroso: 'he who eats safe corn does not develop pellagra, but he who eats rotten corn does'¹⁸. Scores of microorganisms were subsequently tested in various countries in an attempt to find the causative agents.¹³ Two further views developed as outflows of the toxic maize theory, i.e. the belief that good uncontaminated maize itself carries toxic substances, and the view that when spoiled maize is ingested certain toxic substances are formed within the body.¹⁹

In contrast to hypotheses based on the toxic-maize theme, many investigators suspected nutritional deficiencies, rather than toxicity, to be the causal link between maize and pellagra. The Spaniard Gaspar Casal (1650–1759), who first officially documented pellagra, in observing the prevalence of the disease in those subsisting on maize as a staple food, recommended the inclusion of vegetables and cow's milk to their diets. This concept of a nutritional inadequacy was more explicitly defined in 1810 by the Italian Giovanni Marzari who attributed pellagra to an unknown deficiency in diets over-dependent on maize.²⁰ Filippo Lussana (1820–1897), an Italian medical practitioner and clinical neurophysiologist, was more specific, and in a 1856 publication Lussana and Carlo Frua described a proteinaceous insufficiency as the cause of pellagra, declaring the protein content of maize to be far below that required for health.²¹ In 1914, the specific 'nitrogenous substance' referred to by Lussana and Frua, and today known as the substrate for *de novo* synthesis of niacin, was identified as tryptophan by Mendel and Osborne. Lafayette Mendel, a professor of physiological chemistry with an interest in the relative values of various proteins for growth and health, and his co-worker at Yale, Thomas Osborn, a chemist interested in the amino acid composition of plant proteins, demonstrated the low availability of tryptophan and lysine in zein, the major protein of maize.²² During these early years of the study on protein chemistry, several workers were involved in the pursuit of what was deemed 'micronutrients that could prevent disease'. The Polish biochemist Casimir Funk (1884–1967), working successively in Paris, the Lister Institute and the USA, is generally accredited with being the first to use the name vitamins (vitamines) in 1912 and to propose among others an 'antipellagic' vitamin.²³ In a paper, largely ignored at that time, Funk suggested that a change in the method of milling maize was contributory to the prevalence of pellagra.²⁴ As discussed in later paragraphs, this suggestion was, in fact, correct. The so-called 'vitamine hypothesis',

inferred by Funk in early 1912 was later that year articulated by Fleming Sandwith and Rupert Blue, members of the American Clinical and Climatological Association.²⁵

Despite these early observations^{17,19,20}, formal recognition of the disease as a nutritional deficiency only happened at the beginning of the 20th century²⁶. There was some opposition to this recognition, which should largely be seen against the background of the then prevailing knowledge of clinical medicine. The concept of nutritional deficiencies as causes of disease, despite being implied more than a century earlier by individuals such as Italian Giovanni Marzari,²⁰ was in its infancy and the general belief was that good nutrition consisted of proteins, fats and carbohydrates as the essential elements. It is also important to remember that the decisive work that led to confirmation of pellagra as a nutritional deficiency was performed not long after Robert Koch's identification of *Bacillus anthracis* in 1875 – therefore during a period when many diseases of unproven aetiology were suspected of being of bacterial origin. Thus, the commonly held opinion by the early 20th century was that pellagra was an infectious and communicable disease.²⁶ Two theories contributed to the persistence of the germ theory: the spoiled maize theory of Lombroso^{13,18} and Sambon's theory of pellagra as a vector-borne infectious disease^{12,25}.

In 1914, Joseph G. Goldberger confirmed pellagra as a nutritional-deficiency disease by managing to cure, prevent and induce it through dietary manipulations in institutionalised humans.^{26,27} By then, several groups – including the physician Claude Hervey Lavinder, assigned in 1909 to pellagra investigation by the USA Public Health Service – had already shown that pellagra could not be transmitted from humans to other animals.²⁵ Goldberger confirmed that pellagra was not contagious by injecting himself, his wife and his associates with blood from pellagrins. In his quest to demonstrate pellagra as a nutritional deficiency, he showed, in children in an orphanage, that pellagra could be cured by improving their diets. These findings were corroborated when pellagra was induced by inadequate maize-based diets in inmates of the Rankin Prison Farm who participated in the study in return for early release.^{26,27} Goldberger named the causative deficient factor pellagra-preventative-factor, the substance earlier described by Funk as an 'antipellagic vitamine'. The pellagra-preventative-factor was eventually, in 1937, identified as nicotinic acid (niacin) by an American biochemist, Conrad Elvehjem, and his colleagues. This investigation was done in dogs, in which the deficiency presents as black tongue.²⁸ Spies, Blankenhorn and Cooper subsequently showed that niacin also cured pellagra in humans, for which *Time Magazine* dubbed them its 1938 'Men of the Year in Comprehensive Science'.²⁹ The work of Goldberger, Elvehjem, Spies, Blankenhorn and Cooper underlined the role of nutritional deficiencies in disease and paved the way for niacin enrichment of certain foodstuffs. It should be mentioned that similar contributions were made in Italy between 1937 and 1940, in which Frontali, Visco and others demonstrated the therapeutic effects of nicotinic acid. However, as a consequence of the general political situation of the period and publication in Italian and German journals, their papers were less accessible to the international scientific community.¹⁸

While the conundrum surrounding the aetiology of pellagra was, for practical purposes, now resolved, general acceptance of the concept of pellagra as a nutritional deficiency was, particularly in the USA, hampered by two factors. One of these was the findings of the influential US Thompson McFadden Pellagra Commission of Carolina that studied pellagra over the period 1912 to 1914.³⁰ The Commission erroneously concluded that pellagra was an infectious disease and that it was not possible to explain pellagra from the viewpoint of a deficient diet alone.³⁰ The other obstacle was socio-political. The association of pellagra with poverty-linked malnutrition, coupled to the high prevalence in the southern states, presented an embarrassment to southern pride with subsequent public denial of the connotation by politicians.³¹

Spread and prevalence of pellagra

Pellagra in Europe

The earliest records on pellagra appeared during the first half of the 18th century in Spain, but by the second half of the 19th century it had also been described in France, Italy, Romania, and Great Britain, with

incidences of pellagra reported for Portugal, Austria, Bulgaria, Serbia, Croatia, Bosnia, Turkey, the Ionian Island of Corfu and Poland.¹⁵

The symptoms of pellagra were first documented in 1735 in poor peasants of the Asturian region of Spain by the Spanish court physician, Don Gaspar Casal.³² At that stage it was referred to as *mal de la rosa*. Casal not only noted the typical dermatological symptoms, but also the gastrointestinal and neurological features and the association with maize-based diets. His findings were published posthumously in 1762 as a chapter in the *'Historia Natural y Medica de el Principado de Asturias'*.¹⁸ By then, a French physician, Francois Thierly, had, after an earlier visit to Casal, already in 1750 published a description of the disease.^{18,31} The formal naming of the disease as pellagra, in 1771, is ascribed to an Italian physician, Francesco Frapolli.^{18,32} However, the term 'pellagra' – which means 'rough skin' – is derived from the Lombard dialect of Northern Italy and may have been in popular use long before its introduction into medical literature.¹⁸ In 1784, a special hospital for pellagrins and for study of the disease, the Joseph II Pellagra Asylum, was established in Legnano (Milan) by warrant of Emperor Joseph II of the Hapsburgs.¹⁵ The Italian physician Gaetano Strambio (1752–1831), head of the asylum, is said to have written, not only an excellent treatise on the physical symptoms, but also one of the best on the neuropsychiatric manifestations of pellagra.

Pellagra remained endemic for over 200 years in those regions of Italy over-dependent on maize as a staple crop for peasants.^{18,33} In certain areas, up to 3% of agricultural workers were affected. The highest prevalence of pellagra was seen in the North – especially Veneto and Lombardy.³³ For Italy as a whole, the highest prevalence appeared to have been over the period 1871 to 1884 with 104 067 officially reported pellagra cases out of an estimated total population of about 290 107.¹⁵ Records suggest that a progressive decline occurred towards the end of the 19th to the beginning of the 20th century, coinciding with improvements in the socio-economic conditions of the peasantry.^{18,33}

The disease was first reported in France in 1818 in the vicinity of Arcachon in Gascony.¹⁵ Numerous reports appeared during the early 19th century which referred to pellagra among poor peasants in the provinces of Gironde, the Landes, the Haute-Garonne, and areas adjacent to the Pyrenees.³⁴ These reports coincided with a widespread increase in the cultivation of maize, with maize becoming the staple food of poor peasants, especially those in the land tenure system.^{34,35} Hospitals were established at Auch, Montpellier and Pau, for what Marie referred to as 'the terrible pellagrous insanity'.³⁶ Pellagra was a major subject of interest in scientific and sociologic thought and literature from 1829 to 1880 and subsequent social reforms and agricultural changes led to a dramatic decrease in the prevalence of pellagra among the population, with asylums for pellagrins becoming vacant. It is suggested that pellagra was virtually eradicated from large parts of France by 1880.^{15,34,35} The French physician Rousset, who later entered politics, was instrumental in bringing the disease to the attention of physicians, legislators and the French public, and contributed to bringing about legislation to improve the living conditions and diets of the rural poor, to diversify crop cultivation and to increase animal husbandry.^{15,34,35}

The first case in Romania was reported in 1810.^{15,19} By the 1930s, the number of cases had steadily increased until, by 1932, a total of 55 000 cases with 1654 deaths had officially been recorded in the estimated population of about 184 300. Pellagra occurred mainly in poor farm and other labourers for whom about 75% of calories were provided by whole yellow maize. A permanent medical committee was appointed in 1888 to study the aetiology and means of combatting pellagra and many prestigious Romanian doctors and major university centres were involved in the study of pellagra.¹⁹ However, implementation of the findings appears to have been hampered by financial constraints and the fact that authorities were more interested in other diseases.¹⁹

Pellagra apparently never reached epidemic proportions in Great Britain. The first confirmed case was seen in 1863 by Dr James Howden, superintendent of the Montrose Royal Asylum.³⁷ In 1909 a young woman from the Shetlands was admitted to the Royal Edinburgh Asylum and was seen by Drs Brown and Cranston.³⁷ The first officially recorded death

as a result of pellagra in England occurred in 1912 in a boy admitted to St Thomas' Hospital in London with obscure nervous symptoms.³⁸ Three more cases of pellagra in children were described by Box over the period 1913–1914; two of these cases were fatal and, in 1923, two more cases were reported by Hutchison and Paterson.³⁹ A small number of other isolated cases are mentioned in a 1913 paper by Sambon who believed pellagra to have been endemic, if not in Great Britain, than at least in Fifeshire, Forfarshire, Aberdeenshire and the Shetland Islands.³⁷ Although documented evidence does not support the idea that pellagra ever reached endemic proportions, it is possible that many cases may have been overlooked as a result of medical inexperience. This view is supported by papers from a number of doctors in Great Britain. For instance, in 1934, Stannus and Gibson wrote:

*Pellagra is certainly a rare disease in Great Britain: owing to the fact that it is an affection that perhaps very few have learned to distinguish and there can be little doubt that a proportion of the cases go unrecognised.*⁴⁰

By 1941 a similar report appeared in the *British Medical Journal*.⁴¹ The low incidence is inexplicable as at least three major famines occurred in Great Britain over the period of the early history of pellagra: Irish Famine of 1740–1741, Great Irish Famine of 1845–1853 and the Highland Potato Famine of 1846–1856. Malnutrition-related diseases are also indicated by a British Army Medical Service report on the rejection, on medical grounds, of about 40% of volunteers for the Boer War (1899 to 1902).⁴²

Pellagra in Africa

Pellagra has not only been recorded in the normal population in Egypt, but also among prisoners of war (POW) and refugees. Although definitive statistics of the incidence among the general population in Egypt are not available, pellagra is said to have been prevalent in certain country districts with at least 55% of pellagrins being agricultural workers.⁴³ A 1937 report on clinical investigations of inhabitants of villages in the Nile Delta showed a 34% incidence of pellagra. Three-quarters of these pellagrins were subsisting on pure maize bread and parasitic infections of the gastrointestinal tract were almost invariably present. The disease was less common in Upper Egypt, where the diets were also more varied.⁴⁴ According to Miller and Ismail⁴⁵, millet was the staple food in Upper Egypt, while maize was the staple food in Lower Egypt. The first documented institutionalised pellagrins were probably the 40–50 'pellagrous lunatics' reported in 1892.¹⁵ Miller and Ismail's 1920 paper describes a further 757 cases of pellagra admitted to the Khanka (Khānkah) Mental Hospital in Egypt over the period 1912–1920.⁴³ The fatality rate for these patients, whom they referred to as 'insane Egyptian pellagrins', was 32%. Significant numbers of pellagrins were also admitted to the Kasr el Ainy Hospital, as well the Abbassia (Abbāsiyya) Asylum for the Insane in Cairo.⁴⁵ The prevalence of pellagra among POWs and refugees in Egypt was high. During the winters of 1916–1918, the number of pellagrins among the Ottoman POW in Egypt reached epidemic proportions, while much lower rates were observed among the German and Austro-Hungarian prisoners – the difference ostensibly because of the 'European diet' provided to the latter, whereas the Ottoman POWs received a 'non-European diet' of lower nutritional value.^{46,47} Pellagra was also prevalent among the 1916–1917 Armenian refugees at Port Said.⁴⁸

Pellagra is said to have become a problem in South Africa following the outbreak of rinderpest in 1897, which killed most of the cattle, leading to a marked deterioration in diets.⁴⁹ The first quantitative documentation of pellagra in South Africa seems to be in 1906: 150 pellagrins were diagnosed among the 3000 Zulu rebel prisoners, captives of the Bambatha uprising led by Bambatha kaMancinza against British rule and poll tax in KwaZulu-Natal.⁵⁰ It is said that that the prisoners received mainly maize while in captivity, but that inadequate diets before captivity may have contributed.⁵⁰ Based on published reports and their own research, Gillman and Gillman⁵¹ estimated that about 2000 cases of pellagra were documented over the period 1906–1945. In scrutinising the original papers, it is evident that pellagra was predominantly found in black populations with inadequate diets (with maize and maize flour as

the main sources of calories). Beyond the time frame of this paper, but of interest, is the observation that the incidence of pellagra increased until the late 1970s whereafter a decline was seen.

Isolated reports on pellagra in the rest of Africa are available leading up to 1945. According to papers referred to in a 1951 book by Gillman and Gillman⁵¹, two cases were reported in the Gold Coast (Ghana), six cases in Kenya, five cases in Tanganyika (Tanzania), and seven cases in Uganda, while estimates varied from endemic for Rhodesia (Zimbabwe and Zambia) and Sudan to between endemic to epidemic in Nyassaland (Malawi), Basutoland (Lesotho) and Swaziland. In view of periods of starvation as a consequence of natural disasters and population displacements, and in examining later (post-1945) reports on pellagra in Africa, especially among food aid dependent populations, it is feasible to suspect the prevalence to have been much higher than that concluded from official reports.

Pellagra in Asia

The first official records of pellagra in China appear to be in 1923 when six cases were reported in Shanghai and one in Wuhu and, in 1926, with one reported case in Peking (Beijing).^{52,53} Forty more cases of pellagra were reported 14 years later from a hospital for the Shanghai War Refugee Camps.⁵⁴ Wilkinson⁵⁵, in a 1944 publication in *The Lancet*, referred to a high prevalence of nutritional diseases in China, with pellagra having a much higher incidence in the South than in the North. It is possible that the extensive use of soya beans popular in the North, but not in the South, may have accounted for the lower prevalence of pellagra in the North.⁵² Wilkinson blamed pellagra on poverty and a diet of polished rice.⁵⁵ Some support for his suggestion came years later when an outbreak of pellagra (0.57/10 000/day) was seen in March 1994 in Bhutanese refugee camps in Nepal where polished rice formed the emergency food aid.⁵⁶ Suggestions that pellagra was not as rare in China as would be implicated by the sparse available literature are also supported by a limited number of meta-analyses of earlier case histories.⁵⁷

For India, as for China, nothing appears to be known about pellagra before the 20th century. Lowe⁵⁸ reported in 1931 on what may be the first officially recorded cases: 40 pellagrins among inmates of a leper hospital. Over the ensuing decade, isolated cases were reported by Lowe in 1933, Raman in 1933, Manson-Barin in 1935, Rogers and Megaw in 1935, Panja in 1935, Raman and Rau in 1936, Harris in 1939, Gupta et al. in 1939, Bajaj in 1939 and Goodall in 1940,⁵⁹ with most authors suggesting pellagra to be rare in India. Carruthers⁵⁹, who reviewed 10 representative cases from a larger cohort of patients admitted to the Miraj Hospital in Bombay (Mumbai) over a period of 2 years, hypothesised that pellagra was a common condition in the Bombay Deccan and possibly endemic in India; supporting evidence is, however, lacking.

Even less information is freely available on pellagra in trans-continental Russia. Although it is claimed that pellagra was common among prisoners of the Soviet labour camps (the Gulag) that existed between 1923 and 1961,⁶⁰ no official records seem to be available. Nevertheless, from a paper in a Soviet medical journal, describing the eradication of pellagra in the USSR,⁶¹ it is feasible to assume that the disease was prevalent – at least to some extent.

Pellagra in the United States of America

The USA has perhaps the best statistics on the prevalence of pellagra. Mortality data were gathered by each state department of health and compiled by the National Centre for Health Statistics (NCHS) of the US Department of Health and Human Services.⁶² Pellagra is generally said to have first officially been reported in the USA in 1902 in a farmer presenting with dermatological symptoms, severe melancholia and suicidal ideation.²⁶ However, it has been suggested that endemic pellagra was already present after the Civil War, but that the spread of information was suppressed for political reasons.^{31,63} Further evidence that pellagra was known in the USA earlier than 1902 comes from excerpts of the *Transactions of the National Conference on Pellagra*, held under the auspices of the South Carolina State Board of Health at

the State Hospital for the Insane, Columbia, South Carolina, in 1909, in which it was suggested that pellagra had already been discussed at a meeting of asylum superintendents in Washington as far back as 1864.⁶⁴ Furthermore, retrospective analyses showed that practitioners in the Spartanburg area had seen patients with pellagra-like symptoms as early as 1885, and when the case records of the South Carolina Hospital for the Insane were reviewed, it appeared that pellagra-like symptoms had been present among the inmates as early as 1834.²⁶

It is estimated that about 3 million people, out of a reported total population of about 120 million, were affected and 100 000 might have died in the USA over four decades – the majority in the American South.²⁶ This number is believed to be an underestimation as some states may not have reported their pellagra cases.³¹ Pellagra in the USA occurred predominantly in the poor who subsisted on maize as a staple food, such as sharecroppers, tenant farmers, cotton mill workers, residents of orphanages and inmates of state penitentiaries, and was reported, above all, in mental institutions. In 1906, 88 cases of pellagra were cited in the Mount Vernon Hospital for the Coloured Insane with a case fatality rate of 64%. Similar outbreaks were seen over the period 1909 to 1911 in Illinois mental hospitals, with case fatality rates of 50% in Peoria, 40% in Kankakee and 33% at the Elgin State Hospital.⁶⁵ Statistics are available for at least two orphanages – the Methodist Orphanage in Jackson with a pellagra prevalence of 32% and the Baptist Orphanage at Jackson with a prevalence of 60%.³¹

The number of deaths in the USA peaked at about 7000 per year in 1928 with much higher death rates for female than for male individuals and for non-white than for white individuals.⁶⁵ A sharp decline in mortality rates occurred between 1928 and 1932 in the USA, and again from 1939 onwards.^{57,65} Various factors contributed to the decline in the incidence, including a greater awareness of the nutritional values of foods, voluntary bread enrichment by bakers, and food fortification through federal regulation.⁶⁵ Nutrition programmes in the USA were accelerated in the 1940s when it came to the attention of the authorities that 25% of draftees were rejected from military service as a result of evidence of some or other form of malnutrition. However, the more permanent eradication of pellagra occurred only with improvements in the socio-economic status of the workers.

Pellagra and maize

Maize is said to have been domesticated in Mesoamerica around 3500 BC, from where it spread to the American continent and, after European contact in the late 15th and early 16th century, to the rest of the world.⁶⁶ Maize was introduced to Europe in the early 16th century⁶⁷ and, during the 16th and 17th century, largely replaced several other foods as the staple food for the poor and as a cheap source of food for those confined to mental institutions and orphanages in Southern Europe and the USA. Several reasons exist for the preferential cultivation of maize. Maize, when introduced into the Old World, increased the yield of food calories per acre well beyond that previously provided by rye and wheat.³¹ In addition, cultivation was less labour intensive and large areas could be planted more economically – all contributing to maize becoming a cheaper staple crop,³³ affordable to impoverished populations and the institutionalised on inadequate low cost diets. Historically, outbreaks of pellagra were seen, largely, when maize took over as staple food. Whether pellagra made its first appearance only after the introduction of maize into the Old World is perhaps questionable. To quote from a writing dated more than a century ago (1912):

...it is probable that pellagra appeared in Europe long before its scientific description, but was classed as either gastrointestinal, or nervous, or mental, or leprous, or scorbutic diseases. It is, therefore, somewhat difficult to determine whether or not pellagra appeared in Europe before the advent of Indian corn as a food.¹⁵

Pellagra, maize and niacin deficiency

Since confirmation of pellagra as a nutritional deficiency – often caused by an inadequate diet over-dependent on maize – the term pellagra has become synonymous with niacin deficiency. Maize is, however, not totally devoid of niacin. The problem is partially to be found in the bio-availability of niacin in mature maize. Depending on the type of maize, up to 99% of the niacin in hard mature, but not immature, maize is present as niacytin, which is niacin bound up in a complex with hemicellulose that is nutritionally unavailable to humans.⁶⁸ This property is, however, not unique to maize as the niacin in sorghum, wheat, rice and barley has also been shown to be in chemically bound form⁶⁸ – which may, in theory, have contributed to isolated reports of pellagra in populations in which sorghum or rice were consumed⁶⁹. Part of the problem with maize as a staple food appears to reside in the process of food preparation. Boiling of maize is arguably said to not have a significant effect on the availability of niacin, but nixtamalisation (preparation of maize or other grain by soaking and cooking in an alkaline solution), roasting and fermentation of maize have been shown to render the niacin more bioavailable.^{68,70,71} In fact, nixtamalisation, as in the preparation of tortillas, is by many considered the main reason for the apparent absence of pellagra in early pre-Columbian North American, Mesoamerican and South American cultures. However, there are many reasons to doubt this assumption. Although nixtamalisation releases niacin, at least 30%, but probably considerably more, of niacin is destroyed by nixtamalisation as performed by pre-Columbian Mesoamerican cultures.⁷⁰⁻⁷² In addition, other practices and dietary habits known to influence niacin intake also existed. Fermentation of maize was, for instance, performed in the preparation of Pozol (in the form of a dough or drink) which is said to have a higher nutritional quality than either raw maize or tortillas^{70,71}, while huilacoche, a maize fungus rich in amino acids, formed part of the diet of pre-Columbian Mesoamerican culture⁷³. Furthermore, various types of maize exist and indications are that floury maize with a soft, nutritionally more bio-available endosperm, was preferentially used in Mexico, Guatemala and the Andean countries.⁷⁰ However, the most important motive for doubting the assertion of nixtamalisation as the main reason for the ostensible absence of pellagra is that these populations were by no means monophagic populations over-dependent on maize – as evidenced by the many books and papers on Pre-Columbian cuisine.

Beyond the limited bioavailability of niacin in unprocessed mature maize, another, probably even more important, factor that could contribute to the pellagrigenic potential of maize is the amino acid imbalance of zein, the main protein of maize.⁷⁰ Although the total protein quality of common maize is not significantly different from that of most other cereals, zein is deficient in the essential amino acids lysine and tryptophan⁷⁰ – substrates for the *de novo* synthesis of niacin. The tryptophan levels do, however, vary in different types of maize and are influenced by environmental conditions.⁷⁰ In addition, maize contains considerable amounts of the amino acid leucine⁷⁴ which, when present in high concentrations, interferes with *de novo* synthesis of niacin from tryptophan by inhibiting quinolinate phosphoribosyl transferase and, in addition, lowers the body's tryptophan levels⁷. Preliminary indications are that leucine in itself could exacerbate the mental derangements associated with pellagra.⁷⁵

Besides the low bio-availability of niacin and the deficiency of tryptophan in unprocessed maize, these nutrients may be lost in the milling process whereby the aleurone and germ is removed to form the refined product.⁶⁸ Up to 80% of niacin and the two vitamins necessary for *de novo* synthesis of niacin from tryptophan (i.e. riboflavin and pyridoxine) may be removed by milling, not only from maize, but also from rye, wheat, barley, rice and sorghum.⁷⁶ Indications are that degerming may also remove considerable amounts of tryptophan from maize, but not from other cereals.⁷⁶ Crude stone-grinding processes, instead of the newer industrialised milling, has been suggested as the reason for what seems to be a relatively low incidence of pellagra in the early 20th century in South Africa.⁵⁰

Erstwhile competing theories on the pellagra–maize link

From an aetiological point of view, the association between maize and pellagra was predominantly sustained by two competing theories:

Lombroso's spoiled maize theory and the nutritional deficiency theory usually associated with Goldberger. Although the nutritional theory on pellagra is now generally accepted, the spoiled maize theory of Lombroso contributed to the persistence of the association between maize and pellagra even after pellagra was confirmed to be a nutritional-deficiency disease.⁷¹ Not generally acknowledged is the fact that both theories were anticipated by others long before the work of either Lombroso or Goldberger. Front-runners to the work of Goldberger were mentioned in the section on the aetiology of pellagra. Similarly, disease as a result of the ingestion of mould-infested grain was known long before Lombroso's work on microorganisms such as *Aspergillus glaucus*, *Eurotium herbariorum*, *Oidium lactis maidis*, *Penicillium* spp., *Sporisorium maidis* and *Sporotrichum maidis* as potential producers of a pellagra-inducing toxin.^{74,77} It is feasible to accept that Lombroso was familiar with ergotism, which is caused by the ingestion of the alkaloids produced by the *Claviceps purpurea* fungus that infects rye and other cereals. He might also have known about the association between mould-infested maize and pellagra, earlier described by investigators such as Guerreschi and DeRolandis in 1814 and 1824, respectively, and in 1845 by Balardini who suggested *Sporisorium maidis* as the causative fungus.¹³

The idea of an association between pellagra and spoiled maize was not totally abolished by the confirmation of pellagra as a nutritional deficiency. Several continued to study the link, not only with regard to maize, but also with sorghum and other grains, and some even came to consider pellagra as a mycotoxicosis.^{69,72} Various factors contributed to the endurance of Lombroso's spoiled maize theory, not least socio-political motives. Admitting pellagra to be a nutritional deficiency would have reflected poorly on the political and social issues of food availability to the lower social classes in Europe.¹⁸ In the American South, recognition of pellagra as a poverty-associated nutritional deficiency would have been an embarrassment to Southern pride.³¹ The spoiled maize theory on pellagra was also reinforced by evidence of pellagra outbreaks when, under conditions of food emergencies, spoiled maize was imported to areas previously not associated with the disease, such as the outbreak on the Island of Corfu (1858 and 1866), and epidemics in Yucatan, Mexico when maize was shipped under unfavourable conditions of humidity and temperature, first from New York in 1882, and again in 1907 from New Orleans.^{15,67}

Although the concepts of pellagra as a nutritional deficiency and pellagra as the result of spoiled maize are generally seen as two separate theories, it is impossible to overlook potential interactions between them. Spoiled maize, and therefore mycotoxin exposure, is generally more likely to have occurred in populations for which improper methods of food handling and storage are common and malnutrition was already a problem.⁷⁸ Furthermore, toxins from contaminated stored maize could have had devastating effects on agriculture and on livestock, further contributing to poverty and malnutrition,^{78,79} and by implication to the prevalence of pellagra. Of interest is the fact that, although not related to pellagra per se, both the association between spoiled maize and disease, and the association between nutritional deficiencies and disease, have now evolved into major scientific disciplines.

In summary it can be said that the link between maize and pellagra cannot be denied. From an historical point of view, a major reason for the strong association between the disease and maize is that pellagra was often observed when maize, with its limited bio-availability of niacin, as well as its amino acid imbalance and tryptophan deficiency, replaced other more nutritional foodstuffs as staple for peasants and other impoverished populations subsisting on near monophagic diets. More recent evidence of an association between pellagra and the nutritional deficiency of maize, is to be found in the fact that pellagra was until recently still reported in sub-Saharan Africa (Mozambique, Malawi, Zimbabwe, Swaziland) in food aid dependent populations, in food emergencies, in refugee camps and among refugee returnee populations.⁸⁰ Pellagra prevalences of up to 13% were reported and supplementation of the maize-based food aid rations with peanuts, and on occasion niacin, was seen to bring about significant relief.⁸⁰ Enormous efforts went into the early studies on the aetiology of pellagra. Although dermatological symptoms were probably

the major diagnostic indicators during the early history of pellagra, gastrointestinal and neuropsychiatric symptoms were well recognised and diagnosed. Today the diagnosis of pellagra and estimates of niacin deficiency are mainly based on the dermatological symptoms. Niacin is the precursor of the all-important coenzyme NAD and we are starting to understand the mechanisms through which niacin and NAD deficiencies can influence virtually all physiological processes, from cerebral functions to genomic stability. Yet, the levels of niacin and its precursor for *de novo* synthesis (i.e. tryptophan) are still rarely measured for diagnostic purposes or in nutritional surveys.

Conclusions

Pellagra seems to have followed wherever maize became the staple food. This persistent association between maize and pellagra is not unwarranted. It is, however, largely the result of the socio-economic circumstances and near monophagic diets of the underprivileged and the spread of maize, with its limited bioavailable niacin and tryptophan, as a staple food. Largely, two theories sustained the dogged association between maize and pellagra. Although pellagra is now generally accepted as a nutritional deficiency, spoiled maize could have contributed indirectly through its effects on food production.

Acknowledgements

We acknowledge the support of the staff of the Medical Library of the University of Pretoria.

Authors' contributions

M.V. was responsible for the conceptualisation, data collection, writing the initial draft and revisions. P.B. was responsible for data collection and reviewing the writing and revisions. J.L.R. was responsible for critically reviewing the writing and revisions. All authors read and approved the final manuscript.

References

- Hanlon P, Byers M, Wilding JPH, Macdonald HM. Environmental and nutritional factors in disease. In: Walker BR, Colledge NR, Ralston SH, Penman I, editors. Davidson's principles and practice of medicine. London: Churchill Livingstone/Elsevier; 2014. p. 97–132.
- Hegyí J, Schwartz RA, Hegyí V. Pellagra: Dermatitis, and diarrhea. *Int J Dermatol*. 2004;43:1–5. <https://doi.org/10.1111/j.1365-4632.2004.01959.x>
- Bogan KL, Brenner C. Nicotinic acid, nicotinamide and nicotinamide riboside: A molecular evaluation of NAD+ precursor vitamins in human nutrition. *Ann Rev Nutr*. 2008;28:115–130. <https://doi.org/10.1146/annurev.nutr.28.061807.155443>
- National nutrient database for standard reference: USDA ARS [webpage on the Internet]. No date [cited 2018 Apr 11]. Available from: <https://www.ars.usda.gov/northeast-area/beltsville-md-bhnrc/beltsville-human-nutrition-research-center/nutrient-data-laboratory/docs/usda-national-nutrient-database-for-standard-reference/>
- Bipath P, Levay PF, Viljoen M. The kynurenine pathway activities in a sub-Saharan HIV/AIDS population. *BMC Infect Dis*. 2015;15, Art. #346, 12 pages. <https://doi.org/10.1186/s12879-015-1087-5>
- McCormick DB. Two interconnected B vitamins: Riboflavin and pyridoxine. *Physiol Rev*. 1989;69:1170–1198. <https://doi.org/10.1152/physrev.1989.69.4.1170>
- Badawy AA, Lake SL, Dougherty DM. Mechanisms of the pellagrigenic effect of leucine: Stimulation of hepatic tryptophan oxidation by administration of branched-chain amino acids to healthy human volunteers and the role of plasma free tryptophan and total kynurenines. *Int J Tryptophan Res*. 2014;7:23–32. <https://doi.org/10.4137/IJTR.S18231>
- Bilgili SG, Karadag AS, Calka O, Altun F. Isoniazid-induced pellagra. *Cutan Ocul Toxicol*. 2011;30(4):317–319. <https://doi.org/10.3109/15569527.2011.574303>
- Oliveira A, Sanches M, Selores M. Azathioprine-induced pellagra. *J Dermatol*. 2011;38(10):1035–1037. <https://doi.org/10.1111/j.1346-8138.2010.01189.x>
- Hampel JS, Hampel WS. Pellagra and the origin of a myth: Evidence from European literature and folklore. *J R Soc Med*. 1997;90(11):636–639. <https://doi.org/10.1177/014107689709001114>
- Cencelli A. New theories and investigations concerning pellagra. *Lancet*. 1915;185(4781):794–796. [https://doi.org/10.1016/S0140-6736\(01\)64542-2](https://doi.org/10.1016/S0140-6736(01)64542-2)
- Sambon LW. Remarks on the geographical distribution and etiology of pellagra. *Br Med J*. 1905;2(2341):1272–1275.
- Harris HF. Pellagra. New York: Macmillan; 1919. p. 1867–1926.
- History of pellagra. Reynolds-Finley Historical Library [webpage on the Internet]. No date [cited 2016 Aug 12]. Available from: <http://library.uab.edu/locations/reynolds/collections/regional-history/pellagra/history>
- Niles GM. Pellagra, an American problem. *Am J Med Sci*. 1912;144(5):1. <https://doi.org/10.1097/00000441-191211000-00014>
- Gentilcore D. Louis Sambon and the clash of pellagra etiologies in Italy and the United States, 1905–14. *J Hist Med Allied Sci*. 2016;71(1):19–42. <https://doi.org/10.1093/jhmas/jrv002>
- Lombroso C. Studi clinici ed sperimentali sulla natura, causa, terapia della pellagra [Clinical and experimental studies on the nature, cause and therapy of pellagra]. Bologna: Fava e Garagnani; 1869. Available from: <https://archive.org/details/39002086347672.med.yale.edu>. Italian.
- Mariani-Costantini R, Mariani-Costantini A. An outline of the history of pellagra in Italy. *Anthropol Sci*. 2007;85:163–171.
- Andresescu O, Rogozea L. Aetiology theories about pellagra at the end of the nineteenth century and early twentieth century in the vision of Romanian doctors. *Bull Trans Univ Braşov Ser VI Med Sci*. 2013;6(55):93–98.
- Marzari GB. Della Pellagra e della maniera di estirparla in Italia [About pellagra and the manner in which it was eradicated in Italy]. Venice: G. Parolai; 1815. Italian.
- Lussana F, Frua C. Su la pellagra, memoria dei dottori Filippo Lussana e Carlo Frua [Pellagra, memoirs of doctors]. Milan: Coi tipi di G. Bernardoni; 1856. Italian.
- Osborne TB, Mendel LB. The amino-acid minimum for maintenance and growth, as exemplified by further experiments with lysine and tryptophane. *J Biol Chem*. 1916;25:1–12.
- Funk C. The etiology of the deficiency diseases. Beri-beri, polyneuritis in birds, epidemic dropsy, scurvy, experimental scurvy in animals, infantile scurvy, ship beri-beri, pellagra. *J State Med*. 1912;20:341–368.
- Funk C. Studies on pellagra. The influence of the milling of maize on the chemical composition and the nutritive value of maize-meal. *J Physiol*. 1913;47:389–392. <https://doi.org/10.1113/jphysiol.1913.sp001631>
- Bryan CS, Mull SR. Pellagra pre-Goldberger: Rupert Blue, Fleming Sandwith, and the 'Vitamine Hypothesis.' *Trans Am Clin Climatol Assoc*. 2015;126:20–45.
- Rajakumar K. Pellagra in the United States: A historical perspective. *Southern Med J*. 2000;93(3):272–277. <https://doi.org/10.1097/00007611-200093030-00005>
- Goldberger J. The etiology of pellagra, the significance of certain epidemiological observations with respect thereto. *Publ Health Rep*. 1914;29:1683–1686. <https://doi.org/10.2307/4570920>
- Elvehjem CA, Madden RJ, Strong FM, Woolley DW. Relation of nicotinic acid and nicotinic acid amide to canine black tongue. *J Am Chem Soc*. 1937;59:1767–1768. <https://doi.org/10.1021/ja01288a509>
- Sachs RH. White Rose history. Vol I. New York: Exclamation Publishers; 2003.
- Siler JF, Garrison PE, MacNeal WJ. Further studies of the Thompson-McFadden Pellagra Commission – a summary of the second progress report. *J Am Med Assoc*. 1914;63(13):1090–1093. <https://doi.org/10.1001/jama.1914.02570130026006>
- Bollet AJ. Politics and pellagra: The epidemic of pellagra in the US in the early twentieth century. *Yale J Biol Med*. 1992;65:211–221.
- Major RH, Don Gaspar C, Francois T. Pellagra. *Bull Hist Med*. 1944;16:351–361.
- Ginnaio M. La pellagra en Italie à la fin du XIXe siècle : les effets d'une maladie de carence [Pellagra in late nineteenth century Italy: Effects of a deficiency disease]. *Population*. 2011;66:671–698. Italian. <https://doi.org/10.3917/popu.1103.0671>

34. Semba RD. The historical evolution of thought regarding multiple micronutrient nutrition. *J Nutr.* 2012;142(1):143S–156S. <https://doi.org/10.3945/jn.110.137745>
35. Semba RD. Théophile Roussel and the elimination of pellagra from 19th century France. *Nutrition.* 2000;16:231–233. [https://doi.org/10.1016/S0899-9007\(99\)00273-7](https://doi.org/10.1016/S0899-9007(99)00273-7)
36. Marie A. La pellagra. Paris: Giardet Brière; 1908.
37. Sambon LW. The natural history of pellagra. *Br Med J.* 1913;2(2740):1–12.
38. Box CR, Mott FW. Fatal pellagra in two English boys, with the results of the pathological investigation of one case. *Trans R Soc Trop Med Hyg.* 1913;6(5):149–156. [https://doi.org/10.1016/S0035-9203\(13\)90046-8](https://doi.org/10.1016/S0035-9203(13)90046-8)
39. Hutchison R, Paterson D. Pellagra in children in England. *Br Med J.* 1923;2(3276):646–647. <https://doi.org/10.1136/bmj.2.3276.646>
40. Stannus HS, Gibson CR. Pellagra in Great Britain. *QJM New Series.* 1934;3(2):211–236.
41. Robertson DS. Pellagra in Great Britain. *Br Med J.* 1941;1:339. <https://doi.org/10.1136/bmj.1.4182.339>
42. Gratzer W. Terror of the table: The curious history of nutrition. New York: Oxford University Press; 2005.
43. Miller RS, Ismail AC. Pellagra and the pellagra psychosis. A study of 757 insane Egyptian pellagrins. *Lancet.* 1920;196(5068):788–789. [https://doi.org/10.1016/S0140-6736\(01\)19766-7](https://doi.org/10.1016/S0140-6736(01)19766-7)
44. Ellinger P, Hassan A, Taha MM. Pellagra in Egypt. *Lancet.* 1937;230(5952):755–758. [https://doi.org/10.1016/S0140-6736\(00\)88713-9](https://doi.org/10.1016/S0140-6736(00)88713-9)
45. Sandwith FM. The medical diseases of Egypt. London: Kimpton Publishers; 1905. p. 283–310.
46. Yanıkdağ Y. Prisoners of war (Ottoman Empire/Middle East). In: *International Encyclopedia of the First World War.* Berlin: Freie Universität Berlin; 2014. <https://doi.org/10.15463/ie1418.10269>
47. Enright JI. The pellagra outbreak in Egypt: II. Pellagra amongst German prisoners of war: Observations upon the food factor in the disease. *Lancet.* 1920;195(5045):998–1004. [https://doi.org/10.1016/S0140-6736\(01\)08063-1](https://doi.org/10.1016/S0140-6736(01)08063-1)
48. White GR. Report on an outbreak of pellagra in Armenian refugees in Port Said, 1916–1917. In: Ministry of the Interior, Egypt. Cairo: Government Press; 1919.
49. Bender DA. *Encyclopedia of food sciences and nutrition.* 2nd edition. Cambridge, MA: Elsevier; 2003.
50. Cluver EH. Pellagra among the maize-eating natives of the Union of South Africa. *Br Med J.* 1929;2(3590):751. <https://doi.org/10.1136/bmj.2.3590.751-a>
51. Gillman J, Gillman T. Perspectives in human malnutrition. A contribution to the biology of disease from a clinical and pathological study of chronic malnutrition and pellagra. New York: Grune & Stratton; 1951.
52. Lee CU. Pellagra in China: Some recent views on its aetiology. *Nat Med J China.* 1925;11(2):65–74.
53. Jui-Wu M. Pellagra as observed in China: With a report on one case in Peking. *Nat Med J China.* 1927;13(3):229–242.
54. Morris HH, Hwang MS, Kuo PT. Pellagra among the war refugees in Shanghai: Its associated deficiencies and nicotinic acid treatment. *Chinese Med J.* 1940;57(5):427–441.
55. Wilkinson PB. Deficiency diseases in Hong-Kong. *Lancet.* 1944;244(6325):655–658. [https://doi.org/10.1016/S0140-6736\(00\)46017-4](https://doi.org/10.1016/S0140-6736(00)46017-4)
56. Upadhyay J. Persistent micronutrient problems among refugees in Nepal. *Field Exchange.* 1998;5:3.
57. Cui Y, Gao M, Li C, Zhang P, Tang J, Lin Y, et al. Meta-analysis of 23 case reports of pellagra in China. *Pract J Clin Med.* 2009;2:1.
58. Lowe J. Pellagra in India. *Indian Med Gaz.* 1931;66:491.
59. Carruthers L. Pellagra in India. *Trans R Soc Trop Med Hyg.* 1941;35(1):21–34. [https://doi.org/10.1016/S0035-9203\(41\)90067-6](https://doi.org/10.1016/S0035-9203(41)90067-6)
60. Alexopoulos G. The Gulag's veiled mortality [webpage on the Internet]. c2013 [cited 2018 Feb 28]. Available from: www.hoover.org/news/gulags-veiled-mortality-golfo-alexopoulos
61. Spinei LV. The history of eradication of pellagra in the USSR. *Sov Zdravookhr.* 1990;3:65–67.
62. *Vital Statistics of the United States [1900–1978].* DHHS publication PHS 82–1102. Washington DC: National Center for Health Statistics; 1906–1982.
63. Lanska DJ. Stages in the recognition of epidemic pellagra in the United States: 1865–1960. *Neurology.* 1996;47(3):829–834. <https://doi.org/10.1212/WNL.47.3.829>
64. *Transactions of the National Conference on Pellagra; 1909 November 03–04; Columbia, SC, USA.* Columbia, SC: Forgotten Books; 2017.
65. Park YK, Sempos CT, Barton CN, Vanderveen JE, Yetley EA. Effectiveness of food fortification in the United States: The case of pellagra. *Am J Public Health.* 2000;90(5):727–738. <https://doi.org/10.2105/AJPH.90.5.727>
66. *New World Encyclopedia contributors. Maize.* *New World Encyclopedia [encyclopedia on the Internet].* 2014 Sep 10 [cited 2016 Aug 12]. Available from: <http://www.newworldencyclopedia.org/p/index.php?title=Maize&oldid=984265>
67. Roe DA. A plague of corn. The social history of pellagra. Ithaca, NY: Cornell University Press; 1973.
68. Wall JS, Carpenter KJ. Variation in availability of niacin in grain products. Changes in chemical composition during grain development and processing affect the nutritional availability of niacin. *Food Tech.* 1988;10:198–204.
69. Sashidhar RB, Ramakrishna Y, Bhat RV. Pellagra, mycotoxins and tryptophan- niacin metabolism. In: Schwarcz R, Young SN, Brown RR, editors. *Kynurenine and serotonin pathways. Advances in experimental medicine and biology.* Vol 294. Boston, MA: Springer; 1991. p. 631–636. https://doi.org/10.1007/978-1-4684-5952-4_82
70. *Food and Agriculture Organization of the United Nations (FAO). Maize in human nutrition.* FAO Food and Nutrition Series No. 25. Rome: FAO; 1992.
71. Tanumihardjo SDJ. Effects of different processing methods on the micronutrient and phytochemical contents of maize. *Compr Rev Food Sci Food Saf.* 2016;15(5):912–926. <https://doi.org/10.1111/1541-4337.12216>
72. Schoental R. Mouldy grain and the aetiology of pellagra: The role of toxic metabolites of *Fusarium*. *Biochem Rev.* 1980;8:147–150. <https://doi.org/10.1042/bst0080147>
73. Patel S. Nutrition, safety, market status quo appraisal of emerging functional food corn smut (huítlacoche). *Trends Food Sci Technol.* 2016;57(A):93–102. <https://doi.org/10.1016/j.tifs.2016.09.006>
74. Kies C, Fox HM. Interrelationships of leucine with lysine, tryptophan, and niacin as they influence protein value of cereal grains for humans. *Cereal Chem.* 1972;49:223–231.
75. Gopalan C. Possible role for dietary leucine in the pathogenesis of pellagra. *Lancet.* 1969;293(7587):197–199. [https://doi.org/10.1016/S0140-6736\(69\)91206-9](https://doi.org/10.1016/S0140-6736(69)91206-9)
76. Hegedüs M, Pedersen B, Eggum BO. The influence of milling on the nutritive value of flour from cereal grains. 7. Vitamins and tryptophan. *Plant Food Hum Nutr.* 1985;35:175–180. <https://doi.org/10.1007/BF01092134>
77. Lombroso C. *Die Lehre von der Pellagra* (translated into German by H. Kurella) [The study of pellagra]. Berlin: Oscar Coblenz; 1898. German.
78. Bennett JW, Klich M. Mycotoxins. *Clin Microbiol Rev.* 2003;16(3):497–516.
79. Burger HM, Rheeder P. Mycotoxin impact on animal health [webpage on the Internet]. c2017 [cited 2018 Feb 28]. Available from: <http://www.grainsa.co.za/mycotoxin-impact-on-animal-health>
80. Prinzo ZW. Pellagra and its prevention and control in major emergencies. Geneva: World Health Organization; 2000.





Monthly average daily solar radiation simulation in northern KwaZulu-Natal: A physical approach

AUTHOR:
Betty Kibirige¹

AFFILIATION:
¹Department of Physics and Engineering, University of Zululand, KwaDlangezwa, South Africa

CORRESPONDENCE TO:
Betty Kibirige

EMAIL:
kibirigeb@unizulu.ac.za

DATES:
Received: 19 Oct. 2017
Revised: 14 Feb. 2018
Accepted: 22 Feb. 2018
Published: 11 Sep. 2018

KEYWORDS:
Zululand; simulation; Microsoft Excel; model; estimates

HOW TO CITE:
Kibirige B. Monthly average daily solar radiation simulation in northern KwaZulu-Natal: A physical approach. *S Afr J Sci.* 2018;114(9/10), Art. #4452, 8 pages. <https://doi.org/10.17159/sajs.2018/4452>

ARTICLE INCLUDES:
× Supplementary material
× Data set

FUNDING:
University of Zululand

Solar energy is a poorly tapped energy source in northern KwaZulu-Natal (South Africa) and many locations in the region have no available measured solar radiation data. Unfortunately, these areas are among the rural, non-commercial farming areas in South Africa that need to harness solar radiation as an alternative energy source for their needs. These communities are mostly disadvantaged and unable to access the currently sophisticated approaches available for the prediction of such data. For this reason, a modelling tool accessible to these communities has been created using data from the South African Sugarcane Research Institute at eight stations in the region. This article presents the physical approach which can be used within readily available resources such as Microsoft Excel to develop a simulation environment that can predict monthly daily average solar radiation at locations. A preliminary model was later customised by considering the physical condition at each individual location. The validated tool provides estimations with a percentage root mean square error (%RMSE) of less than 1% for all locations except for Nkweleni which had 1.645%. This is an extremely promising estimation process as compared to other methods that achieve estimations with %RMSE of above 10%. The simulation environment developed here is being extended to predict the performance of solar photovoltaic systems in the region. Using data from other sources, the approach is also being extended to other regions in South Africa.

Significance:

- The study modifies the physical approach that is deemed complicated to something that can be accessible to many communities.
- The accuracies achieved with this approach (<0.9%RMSE) in the considered region are commendable.
- The approach can be extended to other regions in South Africa.

Introduction

Northern KwaZulu-Natal is mostly rural with weather data collection stations widely spread apart. The South African Sugarcane Research Institute has data collecting points that leave most areas of interest with no measured records. South African universities have initiated a community that collects weather data and publishes it on the Internet¹ but still these areas are not covered. Some researchers use satellite data to estimate the data at a location. However, these approaches are rather sophisticated and expensive, making them inaccessible for rural communities.

Weather data are used by many disciplines and solar energy conversion is one such field. This field involves engineers, economists, meteorologists, peasant farmers and rural communities to mention a few. Many researchers have estimated solar radiation in areas that have some data collecting stations but these methods are mostly abstract and not easily explicable. Other researchers use satellites to estimate the required data – an approach that is rather sophisticated and expensive for rural communities. This article presents a physical approach using readily available resources like Microsoft Excel to develop an environment that can be used to predict monthly daily average solar radiation and eventually assist in photovoltaic system performance analysis. This approach is a consolidation of various reports on the intensity of solar radiation on the earth's horizontal surface in an effort to come up with a non-abstract prediction of the same in northern KwaZulu-Natal. The model is part of a larger model designed to predict the performance of solar photovoltaic systems in the region. Only the equations deemed to be effective towards the prediction were extracted from the literature. Weather data from eight stations in northern KwaZulu-Natal were used in the analysis of solar radiation trends in the region. Two of the stations with latitudes at the centre of the investigated region were used to generate a preliminary model. Four other stations were used to refine the preliminary model and the final two (with the lowest and the highest latitude) were used to validate the model.

Photovoltaic system designers need to have an estimation of daily monthly average solar radiation for locations where no data collecting stations are present in order to size photovoltaic electrical generating systems appropriately. One of the widely used methods of estimation is a correlation between solar radiation and sunshine hours introduced by Angstrom², and later modified by Prescott³, and the original form of the formula is:

$$\frac{G}{G_0} = a + b \frac{S}{S_0}, \quad \text{Equation 1}$$

where $\frac{G}{G_0}$ is the ratio of the total solar radiation to the available solar radiation at the top of the atmosphere and $\frac{S}{S_0}$ is the ratio of the measured sunshine duration to the theoretically available sunshine hours. The coefficients a and b are calibrated to suit the various location. It is observed from Equation 1 that the total solar radiation G is proportional to the actual effective sunshine hours S of a given location.

Others have estimated mean monthly solar global radiation as a function of temperature.⁴ Here the difference between the maximum and minimum temperatures has been used. These methods include the Bristow–Campbell

model⁵ and the Allen model⁶. Rainfall and temperature measurements have also been used to make the estimation.⁷ Artificial neural networks have been used as an alternative computational approach to estimating daily monthly average global irradiation.^{8,9}

The ethos of these methods is to offer a way of estimating solar radiation without delving into what has been perceived as the more complicated physical approach.¹⁰ These methods are abstract and are not easily explained practically. This study falls back on the more realistic physical approach. With this approach, the prediction of monthly daily average solar radiation averages is done starting with the value of the solar constant, I_0 .¹¹ A simulation environment was developed in Microsoft Excel to assist in the prediction of solar radiation at locations in Northern KwaZulu-Natal, a subtropical region in South Africa. Microsoft Excel is an environment commonly used for computation by many fields and is mostly easily accessible to disadvantaged communities. This tool and other similar open-source tools can be used effectively to make informed decisions on solar energy availability at specified locations. This makes the estimation process open to the user and corrections to the estimations can be made readily if necessary – a welcome advantage that is masked by commercially available software.

Earth's atmosphere and solar irradiance: A review

Photovoltaic cells react to the photo energy incident on their top surfaces to generate a current or voltage that can be harnessed as a source of electrical energy. Photo energy passing through a medium is attenuated according to the properties of the medium. Therefore, photo energy from the sun is scattered by clouds and atmospheric particles, absorbed by atmospheric particles, gases and clouds, diffused and or reflected as it traverses the earth's atmosphere before it makes its way to the surface of the earth. Although most of the radiation received at the surface of the earth is directly radiated from the sun, a substantial amount of it is received from diffused radiation. This is radiation that has been scattered by particles in the atmosphere but still makes its way to the surface of the earth. A conservative estimation is an average of about 10% of the direct radiation, although this varies from dawn up to about 40% at sunset.

The solar irradiance from the sun that is perpendicular to the virtual atmospheric surface of the earth is known as the solar constant I_0 and has a value of approximately 1367 W/m². The optical path through the earth's atmosphere is known as air mass (AM).¹² This mass indicates the path length of the rays of the sun relative to the zenith at sea level. The sea level air mass at the zenith is 1 and increases to 38 as the angle between the rays and the zenith increases to that at the horizon. AM decreases as the altitude above sea level increases and can be less than 1 at some altitudes. AM is simply defined as the ratio of the path length, L , of the rays of the sun through the atmosphere to the surface of the earth to the length, L_0 , if the rays of the sun were directly perpendicular to the surface of the earth (the zenith). Using the zenith angle θ_z between the two lengths, AM is given by the expression:

$$AM = \frac{L}{L_0} = \frac{1}{\cos\theta_z} \quad \text{Equation 2}$$

Photovoltaic modules are normally rated using an AM of 1.5 (AM1.5).¹³

An approximation of the effects of AM on the global solar irradiance G_0 as suggested by Forero et al.¹⁴, Meinel and Meinel¹⁵ quoting Laue¹⁶ and PVEducation¹⁷ for clear skies is:

$$G = G_0 \times 0.7^{AM^{0.678}} \quad \text{Equation 3}$$

modified to include altitude h in km¹⁷:

$$G = G_0 \times [(1 - h/7.1) \times 0.7^{AM^{0.678}} + h/7.1] \quad \text{Equation 4}$$

Equation 4 signifies the solar radiation intensity at the surface of the earth taking into consideration the effects of the earth's atmosphere. For clear weather, especially during the winter solstice in the southern hemisphere, it can be assumed that most of the radiation received at

the earth surface is a result of direct solar radiation. Equation 4 is used as the base equation in the modelling process for this study. Effects of diffusion scatter, absorption and albedo are considered as varying parameters dependent on the time of year and the physical conditions of a location.

Global latitude, longitude and solar irradiance

The equator of the earth is tilted at an angle of 23.45° to the plane of the orbit of the earth around the sun leading to a solar declination angle (δ), defined as the angle between a sun's ray (extended from the centre of the sun to the centre of the earth) and the earth's equatorial plane. This angle varies between 23.45° and -23.45° as the earth orbits the sun and can be expressed as:

$$\delta = -23.45 \times \cos\left(\frac{360}{365} \times (d+10)\right) \quad \text{Equation 5}$$

or

$$\delta = 23.45 \times \sin\left(\frac{360}{365} \times (d+284)\right) \quad \text{Equation 6}$$

The irradiance at the surface of the earth is also affected by the solar elevation angle α . This is defined as the angle between the sun's ray from the centre of the sun to the centre of the earth and the earth's horizontal plane. Daily averages are considered in this study; and the elevation angle is a function of the solar declination δ , and the latitude φ of the location. At noon, in the southern hemisphere, the elevation angle is expressed as:

$$\alpha = 90^\circ + \varphi - \delta \quad \text{Equation 7}$$

The longitude of the location gives an hour shift every 15° longitude and will not affect the solar elevation if daily averages are considered. With the assumption that the earth is spherical, solar radiation intensity at the surface of the earth should vary in a sinusoidal manner to the elevation angle α leading to Equation 8:

$$G' = G \sin \alpha \quad \text{Equation 8}$$

Therefore, apart from diffusion, solar radiation intensity at the surface of the earth varies with the day of the year d , the air mass AM, the altitude of a location h , declination angle δ , the latitude φ and the elevation angle α . The expression is consolidated using Equation 4 and Equation 8 to arrive at Equation 9:

$$G' = (G_0 \times [(1 - h/7.1) \times 0.7^{AM^{0.678}} + h/7.1]) \sin \alpha \quad \text{Equation 9}$$

In the following sections, Equation 9 is translated to read solar radiation intensity I . Factors including the solar constant I_0 , the effective sunshine hours S_{Ave} , and the atmospheric effects (scatter, absorption, diffusion, albedo etc.) are used to estimate the available monthly daily solar radiation.

Sunshine hours

The day length in hours is calculated by considering the theoretical sunrise/sunset angle, ω_s , which is expressed as:

$$\omega_s = \cos^{-1}(-\tan\delta \tan \varphi) \text{ in degrees} \quad \text{Equation 10}$$

At the equator when $\varphi=0^\circ$, the sunrise/sunset angle is 180° and the day length in hours is observed to be 12 h.

The day length in hours S_0 at the equator is derived from Equation 10 by considering that the longitude of the location gives an hour shift every 15° longitude:

$$S_0 = \frac{2}{15} \omega_s \quad \text{Equation 11}$$

S_0 depicts the maximum possible sunshine hours which differ from the effective sunshine hours. The sunshine hours are important in determining the amount of solar energy available in an area during a typical day in a month. Peak sun hours, which are hours during which the intensity of the sun is 1 kW/m^2 , are observed to average about only half of the maximum possible sunshine hours in the region under consideration. An expression derived from Equation 11 above gives these hours, referred to as sunshine hours S , as:

$$S = \frac{\omega_s}{15} \quad \text{Equation 12}$$

A yearly average of the calculated effective sunshine hours S_{Ave} was then used to calculate a solar constant I'_0 in $\text{MJ/m}^2/\text{day}$ using the expression below:

$$I'_0 = \left(\frac{I_0}{11.574} \right) \left(\frac{S_{Ave}}{24} \right) \quad \text{Equation 13}$$

where 24 is indicative of the hours available in a 24-h day (and could be represented by S_0) and 11.574 is a conversion factor from watts (W) to megajoules (MJ). Equation 13 like Equation 1 makes the proportionality of global solar radiation to the actual effective sunshine hours apparent and hence provides a possible explanation of the possible validity of the Angstrom–Prescott equations. That is: $I'_0 \propto \frac{S_{Ave}}{S_0}$ for Equation 13 and $\frac{G_0}{G_0} \propto \frac{S}{S_0}$ for Equation 1.

Absorbed global solar radiation

Clouds and atmospheric particles absorb global solar radiation, which results in a low radiation received at the surface of the earth. Rainy and misty seasons will have higher solar radiation absorption. The effect will be a reduction in the solar radiation, making it necessary to modify Equation 10 with an attenuation factor that is determined in the model development process.

Earth's surface topography and global solar radiation

The study of the shape and features of the surface of the earth – earth's topography – indicates that these features affect the weather of a location.¹⁸ In mountainous terrain, solar radiation measurements may be lower or higher than expected depending on the locality of the measuring site within the terrain. Mountain shadows lower sunshine hours of a location when compared with another location with similar latitude but with a flat terrain.

In essence, the simulated radiation, Rad_{Sim} , can be summarised as in Equation 14:

$$Rad_{Sim} = \beta(I'_0 \times [(1 - h/7.1) \times 0.7^{AM^{0.678}} + h/7.1]) \sin \alpha, \quad \text{Equation 14}$$

where β represents the absorption/diffusion factor depending on the physical implications. This factor takes on a value of 1 if these effects are deemed to be absent at a location.

Methods

Development of the simulation environment

Monthly average daily solar radiation data, collected over long periods of up to 20 years from eight weather stations in the KwaZulu-Natal Province of South Africa,¹⁹ were used for the study.

Figure 1 shows the regional map of the area under study and Table 1 presents the latitudes, longitudes and altitudes together with the period of observation of the chosen sites. In order to establish a way of making estimations for sites without measured data, six of the eight sites were used in the modelling process and the data from the other two sites were used for model validation.



Source: South African Sugarcane Research Institute¹⁹

Figure 1: Regional map of KwaZulu-Natal showing Gingindlovu at the highest latitude and Pongola at the lowest and the positions of the eight weather stations.

Table 1: Sites, their global locations and their period of data collection

Site	Latitude	Longitude	Altitude	Period (years)
Pongola	27°24'S	31°35'E	308 m	20
St Lucia	28°27'S	32°17'E	048 m	12
Mtubatuba	28°27'S	32°13'E	015 m	10
Heatonville	28°43'S	31°48'E	200 m	11
Empangeni	28°43'S	31°53'E	102 m	6
Nkwaleni	28°44'S	31°33'E	118 m	4
Entumeni	28°55'S	31°40'E	546 m	8
Gingindlovu	29°01'S	31°36'E	093 m	5

Two sites – Heatonville and Empangeni – are at the same latitude (28°43'S) and are positioned at a latitude central to that of all the chosen weather stations. These two sites were used to make a preliminary development of the simulation environment. The other four sites (St Lucia, Mtubatuba, Nkwaleni and Entumeni) were used to refine the model further. Pongola and Gingindlovu were chosen for the validation of the simulation environment. These stations are positioned at the lowest (27°24'S) and the highest (29°01'S) latitudes of the region under study, respectively.

The development of the simulation environment was carried out in Microsoft Excel. A field labelled d was created as the variable for each mid-month day (about the 15th of each month). Then other fields including latitude (φ), declination angle (δ) (Equation 7); solar constant (I'_0) (Equation 13); a factor caused by air mass and altitude ($AM \& h$) (Equation 15); simulated radiation (Rad_{Sim}) and collected radiation data (Rad_{data}) were created as shown in Table 2. Table 2 represents the proposed simulation environment that can be developed for any location if latitude, longitude, altitude and other geographical variables are known for the location.

$$AM \& h = ((1 - h/7.1) \times 0.7^{AM^{0.678}} + h/7.1) \quad \text{Equation 15}$$

In calculating the solar constant in $\text{MJ/m}^2/\text{day}$ from 1367 W/m^2 , an average value of 6 sunshine hours per day was used as is appropriate for the region. This was determined by using Equation 12 for each middle day of the month to determine S_m and then a yearly average S_{Ave} calculated in Excel, for 12 months of the year. The subscript m stands for month.

$$S_{Ave} = \frac{\sum_{m=1}^{12} S_m}{12} \quad \text{Equation 16}$$

Table 2: Example of fields created in Microsoft Excel for each site simulation

Heatonville simulation								
Month	<i>d</i>	φ (degrees)	δ (degrees)	I'_0 (MJ/m ² /day)	<i>AM & h</i>	α (degrees)	Rad_{sim} (MJ/m ² /day)	Rad_{data} (MJ/m ² /day)
Jan	15	28.71667	21.31008	29.52739	0.608387	1.699204	19.59781	19.4
Feb	46	28.71667	13.36471	29.52739	0.608387	1.837806	19.06026	18.8
Mar	75	28.71667	2.504361	29.52739	0.608387	2.027259	17.73736	17.2
Apr	106	28.71667	-9.70854	29.52739	0.608387	2.240306	14.0861	14.1
May	136	28.71667	-18.9854	29.52739	0.608387	2.402136	12.10576	12
Jun	166	28.71667	-23.3065	29.52739	0.608387	2.477516	10.51826	10.3
Jul	196	28.71667	-21.5441	29.52739	0.608387	2.446771	11.50149	11.1
Aug	227	28.71667	-13.834	29.52739	0.608387	2.312273	13.24801	13
Sep	258	28.71667	-2.274	29.52739	0.608387	2.110615	15.40963	15
Oct	288	28.71667	9.551474	29.52739	0.608387	1.904326	16.12543	15.9
Nov	319	28.71667	19.12038	29.52739	0.608387	1.737402	17.71534	17.8
Dec	349	28.71667	23.33097	29.52739	0.608387	1.663951	19.67482	20.1

Graphical representations of simulated and collected data of monthly average daily solar radiation for each site were plotted and the process of fitting simulated data to collected data was carried out.

Analysis

Analysis of results was done using statistical methods. The percentage mean bias error (%MBE), the %RMSE and the coefficient of determination (R^2) were used. Where:

$$\%MBE = \frac{100 \times \sum Rad_{sim} - Rad_{data}}{\sum Rad_{data}} \% \quad \text{Equation 17}$$

$$\%RMSE = \frac{100 \times \sqrt{\sum (Rad_{sim} - Rad_{data})^2}}{\sum Rad_{data}} \% \quad \text{Equation 18}$$

The %RMSE method in Equation 18 gives a true reflection of the data fit. By squaring the differences between simulated and collected data, and then determining the square root of the squared differences, negative differences and positive differences are allowed to contribute to the solution without one cancelling out the other, hence giving a realistic fit.

The %RMSE and %MBE show the best fit when the values tend towards zero while an R^2 is indicative of a good fit when its value tends towards 1.

Results and discussion

Figure 2 shows simulated and collected data before and after including the diffusion factor. The fitted linear regression is shown, together with the corresponding coefficient of determination, R^2 , for Heatonville and Empangeni. Data from Heatonville and Empangeni weather stations were used to start the process of fitting simulated data to collected data.

It is observed from the preliminary data (Figure 2a), that on days on which simulated and collected data did not agree, the simulated data had a lower value than the collected data. The days on which this happens are representatives of the summer solstice months (December to March) in the southern hemisphere.

Seasonal changes carry with them varying atmospheric conditions. The summer months in the studied region are prone to have many particles in the atmosphere that scatter the direct radiation. This scattered radiation still makes its way to the surface of the earth, resulting in diffusion. By including the diffusion factor introduced in Equation 4, a 1.1 factor estimated by Meniel and Meniel¹⁵ and Laue¹⁶, fitted simulated data to

collected data for the months of January and March; but a factor of 1.2 was needed to fit the data for the months of December and February at both stations.

As discussed, clouds and atmospheric particles scatter but also absorb global solar radiation. The value of the simulated data in the months of June and October is higher than that of the collected data. As recorded by the South African Sugarcane Research Institute¹⁹, October and June are rainy months but also colder than the summer months, leading to absorption rather than diffusion. A factor of 0.95 was needed to fit the two sets of data. After including these changes, the simulated and collected data were as shown in Figure 2b. Figure 2c shows a comparison between collected data and simulated data. The fitted linear regression and the corresponding R^2 are 0.9946 and 0.9931 for Heatonville and Empangeni, respectively. The %MBE and the %RMSE are shown in Table 3. The preliminary simulation environment showed a very good correlation between the two data sets and was further used on the four other locations.

Table 3: Performance of the preliminary simulation environment

Site	R^2	%MBE	%RMSE	Period (years)
Heatonville	0.9946	1.126	0.563	11
Empangeni	0.9931	0.423	0.578	06

Results for simulations carried out for the four other stations using the preliminary model developed for Heatonville and Empangeni are presented in Figure 3 and Table 4.

A summary of the results of the stations after using the preliminary Heatonville/Empangeni model is given in the before column given in Table 4. Although the simulations for St Lucia and Mtubatuba showed good correlation with %RMSEs as low as 0.886% and 0.726%, respectively, those for Nkweleni and Entumeni did not. For both of these stations, the %RMSE was as high as 3%.

Better data fitting at Nkweleni was achieved by changing the value of the sunshine hours from the 6 h used as the average for the region to 5.5 h. Figure 4 shows the graph after changing the value of the sunshine hours.

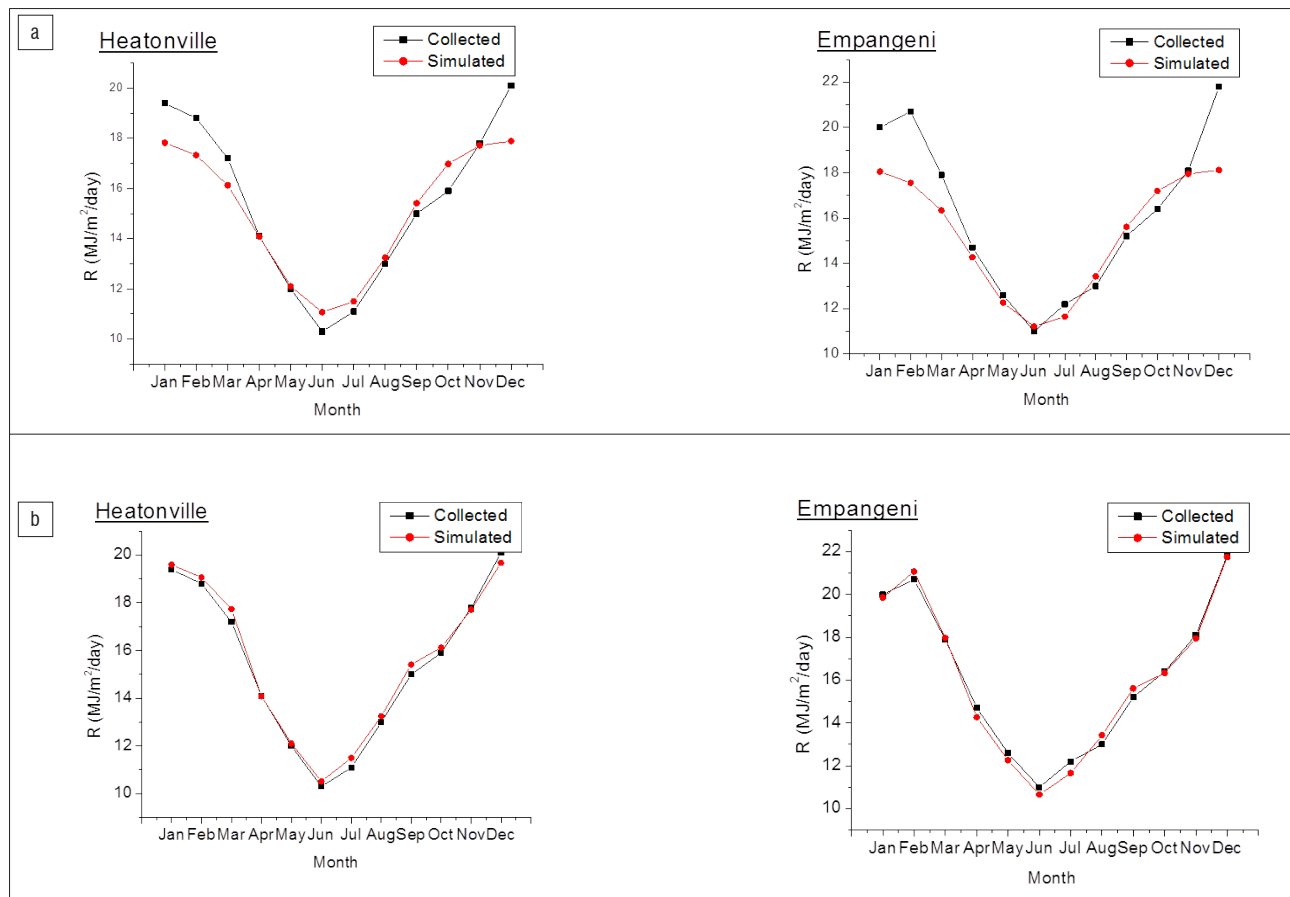


Figure 2: Simulated and collected data (a) before and (b) after inclusion of the diffusion factor for Heatonville and Empangeni.

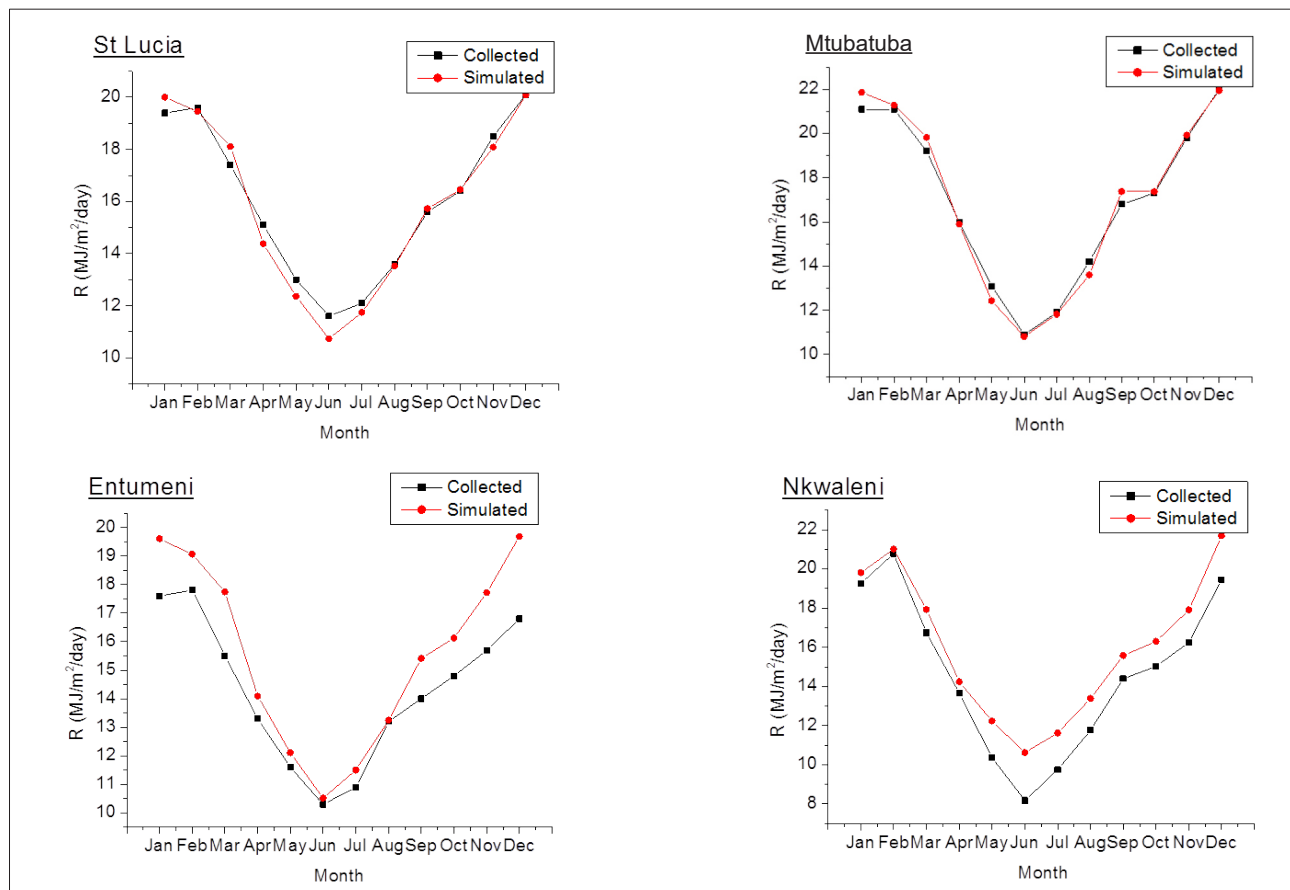


Figure 3: Testing the Heatonville/Empangeni model using four other stations in the studied region.

Table 4: Performance of the simulation environment after correction for four other sites

Site	R^2		%MBE		%RMSE		Period (years)
	Before	After	Before	After	Before	After	
St Lucia	0.985	0.985	-0.912	-0.912	0.886	0.886	12
Mtubatuba	0.992	0.992	0.365	0.365	0.726	0.726	10
Nkwaleni	0.977	0.977	9.517	0.391	3.040	1.654	4
Entumeni	0.971	0.975	8.909	0.223	3.082	0.845	8

Further investigations reveal that Nkwaleni is surrounded by mountainous terrain resulting in shadows over the station and hence reduced sunshine hours at the station. Therefore, mountainous terrain could lead to more exposure to the sun, necessitating an increase in the averaged sunshine hours; or, the same physical feature could lead to shadows, necessitating a decrease in sunshine hours.

Data captured from Entumeni weather station did not agree with the simulated data for the months of November, December, January and March. The data simulated using the Heatonville/Empangeni model, for the location, were found to be higher than the collected data at the location. Entumeni is a misty area, especially in the summer months, which leads to a reduction in the received solar radiation in those months. Including an absorption factor of 0.95 for these months, fitted the data within a desirable range. Figure 4a shows the corrected simulation data.

The results of the analysis of the four stations are summarised in Table 4.

Further validation, results and discussions

This study showed that the daily monthly average radiation received at any location will be affected by: latitude, altitude, day of the year, the earth terrain at the location and AM. Pongola and Gingindlovu were used to validate the model further. Data at these stations were estimated by considering the physical condition and the global data given in Table 1. An estimation of the expected solar radiation was made starting with the Heatonville/Empangeni model.

Gingindlovu

At a latitude of 29°01'S and an altitude of 93 m, Gingindlovu is 7 km from the coast and has a relatively flat terrain with no extreme weather patterns. It was therefore deemed sufficient to rely on the Heatonville / Empangeni model without any further correction.

The simulated and collected weather data were plotted on the same graph and a comparison made between the two sets of data. Results are shown in Figure 5. A summary of the model performance is given in Table 5. The model performed well with a %RMSE of 0.91.

Pongola

At a latitude of 27°24'S and an altitude of 308 m, Pongola is situated on a mountainous slope prone to receive more sunshine in the day. As the sunshine hours of Nkwaleni, a location situated in a shadow, were reduced from 6 h to 5.5 h, the sunshine hours for Pongola were increased from the averaged 6 h to 6.5 h to account for perceived extra sunshine hours.

The simulated and collected radiation data are presented in Figure 5. Table 5 gives a summary of the results for Gingindlovu and Pongola. A model performance and a %RMSE of 0.894 was achieved for Pongola.

Table 5: Performance of the preliminary simulation environment for Gingindlovu and Pongola

Site	R^2	%MBE	%RMSE	Period (years)
Gingindlovu	0.9887	1.328	0.910	5
Pongola	0.9989	0.841	0.894	20

Summary and conclusion

A physical approach has been used to estimate the monthly average daily solar radiation of a demarcated region in northern KwaZulu-Natal, South Africa. Starting with the solar constant, various mathematical expressions that represent the physical processes that the solar radiation encounters on its way to the earth's surface were drawn upon to consolidate a complete model that provides an estimation of the daily monthly average solar radiation at a location within a region. This approach reduces the complications of the physical approach by including only equations that are significant to the modelling process. Latitude, altitude, day of the year, and the location's geographical features were used to develop a simulation environment in Microsoft Excel that gave a prediction of radiation for eight locations, two of which (Gingindlovu and Pongola) were used in validating the model. A preliminary model was developed using two of the locations and the other four locations were used to establish ways of customising the preliminary model to suit a given location. It is noted that mountainous regions could either lead to an increase or a reduction in sunshine hours. This trend was brought out clearly using the developed model to predict sunshine radiation at two of the stations (Nkwaleni and Pongola). Entumeni is a misty place and the solar intensity at the station was observed to be lower than the preliminary estimations, implying a reduction as a result of absorption by the particles in the atmosphere. A factor that compensated for diffused radiation was added in the summer months which confirms that diffused radiation is most prevalent in the summer solstice.

The model gave good results for most months at the considered stations. Solar radiation at Gingindlovu and Pongola was predicted at less than 1%RMSE for both locations and for all other locations except for Nkwaleni which was 1.645%RMSE. This is a very good fit and this approach could therefore be used, with minimal customisation, to estimate energy yields from solar systems around northern KwaZulu-Natal region. The Microsoft Excel program used to carry out the estimations is readily available and can be used easily by many communities. This approach could be used for other regions that have similar needs.

Future work

The simulation environment developed here is being extended to include the prediction of a system's performance for the same region in South Africa. While assisting communities in rural Zululand has been a major motivation for this paper, working with overseas collaborators has highlighted the need for location-specific simulations in order to minimise discrepancies in systems' predictions.

Other data sources like the South African Universities Radiometric Network (SAURAN) do exist. Therefore, as a future project, the approach will be extended to other regions in the country using other data sources.

Acknowledgements

The University of Zululand Research Office is acknowledged for providing funding.

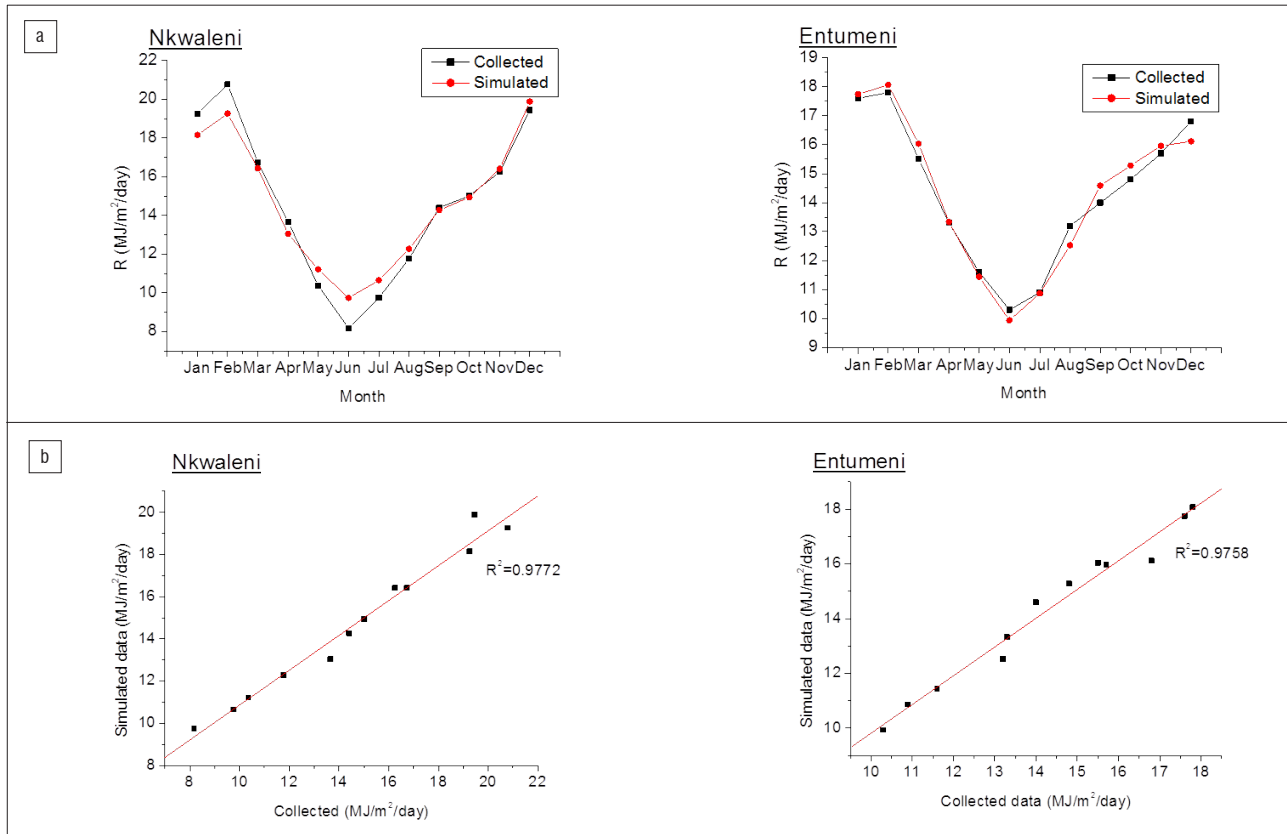


Figure 4: (a) Corrected simulation results for Nkwaleni and Entumeni. (b) The fitted linear regression is shown together with the corresponding coefficient of determination (R^2) for Nkwaleni and Entumeni.

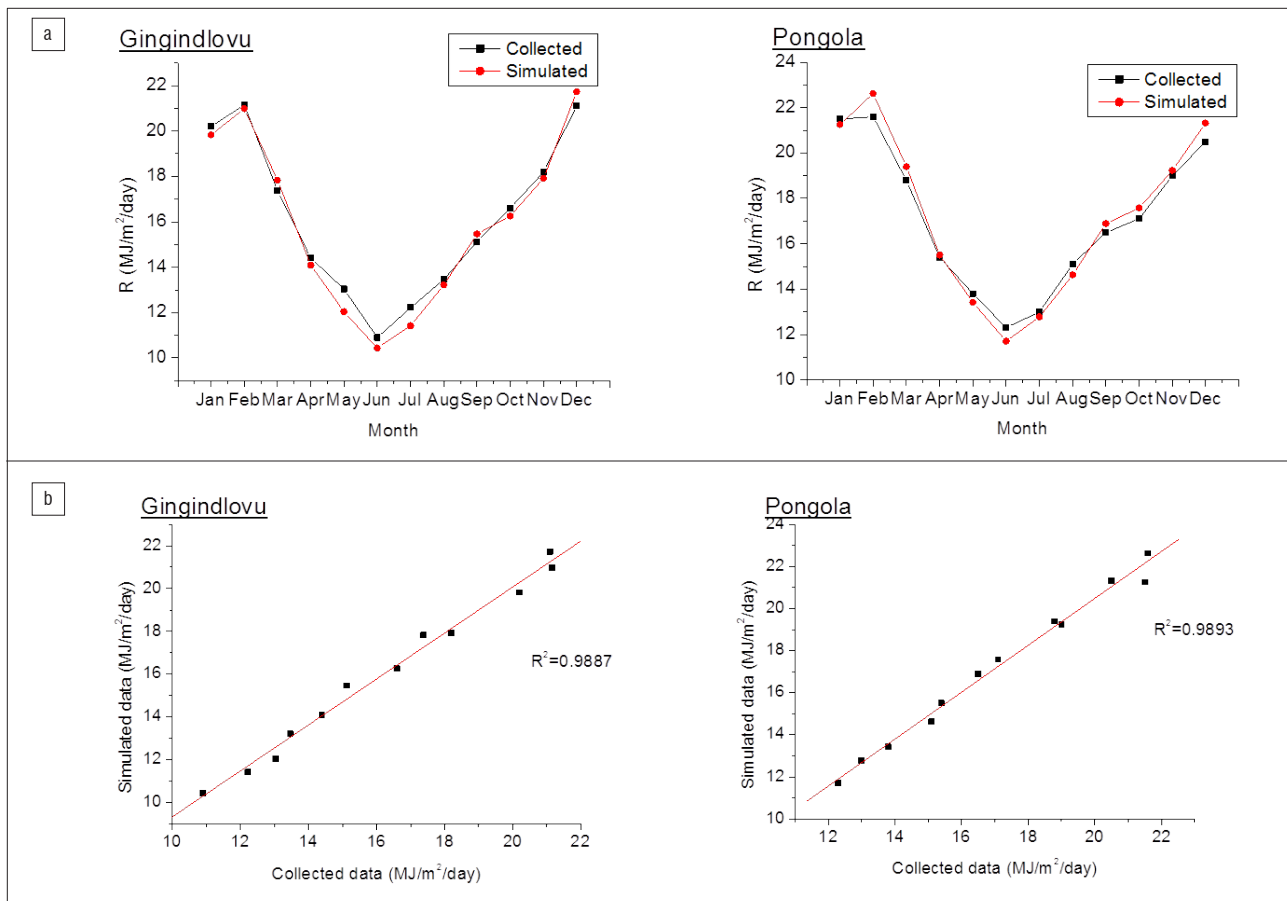
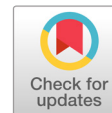


Figure 5: (a) Estimated radiation data for Gingindlovu and Pongola. (b) The fitted linear regression is shown together with the corresponding coefficient of determination (R^2) for Gingindlovu and Pongola.

References

1. Southern African Universities Radiometric Network. Solar radiometric data for the public [webpage on the Internet]. No date [cited 2018 Feb 14]. Available from: <http://www.sauran.net>
2. Angstrom A. Solar and terrestrial radiation. *Q J Roy Met Soc.* 1924;4:121–126.
3. Prescott JA. Evaporation from a water surface in relation to solar radiation. *Trans Roy Soc S Aus.* 1940;64:114–125.
4. Meza F, Varas E. Estimation of mean monthly solar global radiation as a function of temperature. *Agric For Meteorol.* 2000;100:231–241. [https://doi.org/10.1016/S0168-1923\(99\)00090-8](https://doi.org/10.1016/S0168-1923(99)00090-8)
5. Bristow K, Campbell G. On the relationship between incoming solar radiation and daily maximum and minimum temperature. *Agric For Meteorol.* 1984;31:159–166. [https://doi.org/10.1016/0168-1923\(84\)90017-0](https://doi.org/10.1016/0168-1923(84)90017-0)
6. Allen R. Self-calibrating method for estimating solar radiation from air temperature. *J Hydrol Eng.* 1997;2:56–67. [https://doi.org/10.1061/\(ASCE\)1084-0699\(1997\)2:2\(56\)](https://doi.org/10.1061/(ASCE)1084-0699(1997)2:2(56))
7. Liu DL, Scott BJ. Estimation of solar radiation in Australia from rainfall and temperature observations. *Agric For Meteorol.* 2001;106:41–59. [https://doi.org/10.1016/S0168-1923\(00\)00173-8](https://doi.org/10.1016/S0168-1923(00)00173-8)
8. Mubiru J, Banda EJKB. Estimation of monthly average daily global solar irradiation using artificial neural networks. *Sol Energy.* 2008;82:181–187. <https://doi.org/10.1016/j.solener.2007.06.003>
9. Tymvios FS, Jacovides CP, Michaelides SC, Scouteli C. Comparative study of Angstrom's and artificial neural networks methodologies in estimating global solar radiation. *Sol Energy.* 2005;78:752–762. <https://doi.org/10.1016/j.solener.2004.09.007>
10. Dewitte S, Crommelynck D, Mekaoui S, Joukoff A. Measurement and uncertainty of the long term total solar irradiance trend [document on the Internet]. c2004 [cited 2018 Feb 14]. Available from: www.kmi.be/meteo/download/de/519987/pdf/rmi_scpub-898.pdf
11. Willett HC. American air mass properties. Cambridge, MA: Massachusetts Institute of Technology and Woods Hole Oceanographic Institution; 1933. <https://doi.org/10.1575/1912/1142>
12. American Society for Testing and Materials (ASTM). Terrestrial reference spectra for photovoltaic performance evaluation. ASTM G173-03. West Conshohocken, PA: ASTM; 2012.
13. Salazar GA, Hernández AL, Saravia LR. Practical models to estimate horizontal irradiance in clear sky conditions: Preliminary results. *Renew Energy.* 2010;35:2452–2460. <https://doi.org/10.1016/j.renene.2010.01.033>
14. Forero N, Caicedo LM, Gordillo G. Correlation of global solar radiation values estimated and measured on an inclined surface for clear days in Bogota. *Renew Energy.* 2007;32:2590–2602. <https://doi.org/10.1016/j.renene.2006.12.012>
15. Meinel AB, Meinel MP. Applied solar energy, an introduction. Reading, MA: Addison-Wesley; 1976.
16. Laue EG. The measurement of solar spectral irradiance at different terrestrial elevations. *Sol Energy.* 1970;13:43–50. [https://doi.org/10.1016/0038-092X\(70\)90006-X](https://doi.org/10.1016/0038-092X(70)90006-X)
17. PVEducation.org. Welcome to the Photovoltaic Education Network [webpage on the Internet]. No date [cited 2017 Jul 10]. Available from: <http://www.pveducation.org/pvc/drom/2-properties-sunlight/air-mass>
18. Aguilar C, Herrero J, Polo MJ. Topographic effects on solar radiation distribution in mountainous watersheds and their influence on reference evapotranspiration estimates at watershed scale. *Hydrol Earth Syst Sci.* 2010;14:2479–2494. <https://doi.org/10.5194/hess-14-2479-2010>
19. South African Sugarcane Research Institute. Unlocking the potential of sugarcane [webpage on the Internet]. No date [cited 2016 Dec]. Available from: <http://www.sasa.org.za/sasri.aspx>





AUTHORS:

Makandwe Nyirenda¹
Renee Street^{1,2} 
Tarylee Reddy³
Susie Hoffman^{4,5}
Suraya Dawad^{1,6}
Kelly Blanchard⁷
Theresa M. Exner⁸
Elizabeth A. Kelvin^{6,9}
Joanne E. Mantell⁵
Gita Ramjee¹

AFFILIATIONS:

¹HIV Prevention Research Unit, South African Medical Research Council, Durban, South Africa

²Environment and Health Research Unit, South African Medical Research Council, Durban, South Africa

³Biostatistics Unit, South African Medical Research Council, Durban, South Africa

⁴Department of Epidemiology, Columbia University, New York, New York, USA

⁵HIV Center for Clinical and Behavioural Studies, New York State Psychiatric Institute and Columbia University, New York, New York, USA

⁶National Department of Health, Pretoria, South Africa

⁷Ibis Reproductive Health, Cambridge, Massachusetts, USA

⁸Department of Epidemiology and Biostatistics, City University of New York, New York, New York, USA

⁹CUNY Institute for Implementation Science in Population Health, City University of New York, New York, New York, USA

CORRESPONDENCE TO:

Makandwe Nyirenda

EMAIL:

makandwe.nyirenda@mrc.ac.za

DATES:

Received: 14 Nov. 2017

Revised: 07 Feb. 2018

Accepted: 05 May 2018

Published: 11 Sep. 2018

KEYWORDS:

household food insecurity; HIV counselling; HIV testing; primary health care; South Africa

Food insecurity, HIV status and prior testing at South African primary healthcare clinics

HIV and food insecurity are two prominent causes of morbidity and mortality in sub-Saharan Africa. Food insecurity has been associated with risky sexual practices and poor access to healthcare services. We describe the association between household food insecurity and previous HIV testing and HIV status. We used logistic regression to analyse the association between food insecurity and prior HIV counselling and testing (HCT) and testing HIV positive. A total of 2742 adults who presented for HCT at three primary healthcare clinics in KwaZulu-Natal, South Africa, participated in the study. The prevalence of household food insecurity was 35%. The prevalence of food insecurity was highest in adults who had incomplete high schooling (43%), were unemployed (39%), and whose primary source of income was government grants (50%). Individuals who were food insecure had significantly higher odds of testing HIV positive (adjusted odds ratio 1.41, 95% CI 1.16–1.71), adjusted for demographic and socio-economic variables. There was no association between food insecurity and prior HCT. The findings of this study highlight the important role food insecurity may play in HIV risk. Interventions to turn food-insecure into food-secure households are needed to reduce their household members' vulnerability to HIV acquisition. The absence of such interventions is likely to severely impact ambitious global targets of ending AIDS by 2030 through the 90-90-90 targets and test-and-treat-all initiatives.

Significance:

- One in three adults presenting for HIV counselling and testing came from households with some degree of food insufficiency.
- Experience of food insecurity was very high in young people who did not complete high school and were currently not studying.
- Findings support the need for socio-economic and structural interventions to transform food-insecure into food-secure households.
- Failure or lack of such interventions will contribute to the failure to achieve global targets like the UNAIDS 90-90-90 programme.

Introduction

To end AIDS as a global pandemic by 2030, UNAIDS has set the world ambitious targets of 90-90-90.¹ That is, by 2020, 90% of all persons living with HIV will be tested and know their status; of those tested positive, 90% will be on antiretroviral treatment; and finally, 90% of those on antiretroviral treatment will achieve viral suppression. HIV counselling and testing (HCT) is thus the initial step in the cascade of HIV prevention, care and treatment towards ending HIV as a public health pandemic.^{1,2} A number of factors have been identified as contributory to the low uptake of HIV testing services including hunger, poor nutrition and food insecurity.³ In South Africa, the recent National Health and Nutrition Examination Survey revealed that 28% of the population are at risk of hunger and 26% are household food insecure (ever experienced hunger).⁴

Much of the empirical research focus has been on the relationship between food insecurity and the health and well-being of HIV-infected persons.⁵⁻⁹ Food insecurity has been shown to be associated with poor health and poor adherence to medication, including antiretroviral drugs.¹⁰⁻¹² In South Africa, one of the main reasons cited for not accepting free antiretroviral drugs was fear of taking medication on an empty stomach as a consequence of food insufficiency.^{13,14} In Tanzania, supplementary food cost was cited as an access challenge in sustaining long-term antiretroviral treatment.¹⁵

Food insecurity also has been suggested to be a major contributor to risky sexual behaviour and HIV spread in neighbouring Botswana and Swaziland¹⁶ and elsewhere. Those who report food insecurity have been found to be less likely to use condoms than those who report food security.^{17,18}

Much less research has focused on the effect of food insecurity on health-seeking behaviour including HIV testing. To date, there is no information on the role of household food insecurity in HIV testing history and/or testing HIV positive among South Africans.

To address this gap, the aim of this study was to determine the association of household food insecurity with prior HIV testing and HIV serostatus among individuals presenting for HCT at public primary healthcare clinics in KwaZulu-Natal, South Africa. We hypothesised that household food insecurity would be associated with lower likelihood to have previously tested for HIV and higher likelihood to be HIV positive when tested.

Methods

Study population

Data for these cross-sectional analyses were drawn from interviews conducted between November 2010 and May 2012 as part of the Pathways to Care study, a prospective cohort study of newly diagnosed HIV-positive

HOW TO CITE:

Nyirenda M, Street R, Reddy T, Hoffman S, Dawad S, Blanchard K, et al. Food insecurity, HIV status and prior testing at South African primary healthcare clinics. *S Afr J Sci.* 2018;114(9/10), Art. #4407, 6 pages. <https://doi.org/10.17159/sajs.2018/4407>

ARTICLE INCLUDES:

- × Supplementary material
- × Data set

FUNDING:

US National Institute of Mental Health; South African Medical Research Council

women and men, described elsewhere.¹⁹ Women and men presenting for HCT at three public-sector primary healthcare clinics in the Durban region (KwaZulu-Natal, South Africa) were invited to be interviewed prior to HCT. Eligibility for the interview included being 18 years or older, anticipating residing in the community for at least 1 year, being not pregnant, able to speak and understand English or isiZulu, and willing to have HIV test results shared with the study interviewer. Of the population presenting for HCT at the three primary healthcare clinics, 2996 individuals were approached to participate in the study. Of those approached, 29 were not interested and 76 were ineligible based on: age ($n=30$), pregnancy ($n=31$), would not be residing in the study community for longer than a year ($n=8$), and cognitive impairment ($n=7$). A further 106 individuals who had previously tested HIV positive, 41 individuals with incomplete pre-screening or screening assessments, and 2 individuals with missing data on key study variables were excluded, leaving a final sample size of 2742 individuals for this analysis. For the analysis of HIV status, there were 2693 individuals, as 5 were not tested and it was not possible to determine results for 44 individuals. Written informed consent was obtained prior to the screening interview, and the principles of the Declaration of Helsinki were adhered to throughout the study. Participants were reimbursed ZAR40 (the approximate equivalent of USD5 at the time) for the interview. All study procedures were approved prior to study initiation by the University of KwaZulu-Natal Biomedical Research Ethics Committee and the NYS Psychiatric Institute of the Columbia University Institutional Review Boards.

Data sources

A structured questionnaire of approximately 20-min duration was administered by trained interviewers. Items included socio-demographic factors, history of prior HIV testing, reasons for testing, and who recommended testing. HCT was conducted by clinic staff according to standard rapid testing procedures, after which participants returned to the study interviewer with a card that indicated their HIV test result using a code to blind the information. If participants did not return to the study interviewer after HCT, the interviewer obtained test results directly from the clinic testing counsellor.

Measures

Outcome variables

Positive HIV test: This was a dichotomous variable indicating whether the person tested positive or negative on the day of their visit to the primary healthcare clinic.

Previous HIV testing: This variable was categorised as 'never' if the individual had no history of previous testing for HIV or 'ever' if the individual had ever taken an HIV test before presenting for HCT at the health facility.

Predictor

Food insecurity: The dichotomous household food insecurity variable (ever/never) was created from responses to the question 'Do the people in your household go without food?' The 'ever' category consisted of response options 'often', 'sometimes' or 'seldom', and 'never' was for those who never went without food in their household.

Confounders

Informed by available empirical literature,^{17,20,21} potential confounders controlled for in the analysis of HIV status included age categorised into age ranges (<24, 25–28, 29–31, 32–36, ≥37 years), gender, clinic site, educational attainment (completed high school or not), relationship status (not married or in a relationship; married or in a relationship, not living together; married or in a relationship, and living together), employment status (employed part-time/full-time/self-employed or unemployed), primary source of income (none, family, employment, spouse/partner, government grant, or other), caring for dependent adults and/or children (none, children only, adults only, both), ability to borrow money (up to ZAR100) if needed for medical expenses (yes/no), and time taken to travel to the clinic (<30 min, 30–60 min, >60 min).

Statistical analysis

Chi-square tests were used to evaluate the association between each demographic and socio-economic factor and household food insecurity. The relationship between household food insecurity and previous HIV testing and HIV status was examined using univariable and multivariable logistic regression models. A variable was deemed a potential confounder if its inclusion in the model containing food security resulted in a 10% or higher change in the model coefficient for food security.

Results

The majority of the participants in this analysis were women (52.8%), were aged less than 24 years (36.5%), had not completed high school (61.5%), were not currently studying (87.4%) and were married or in a relationship but not living together (62.6%) (Table 1). Further characteristics of the sample were that most were unemployed (65.1%) but were currently caring for a dependent adult or child (58.5%). Nearly two-thirds of participants came from households with at least one government grant recipient (63.1%). Nearly all participants (96.4%) travelled less than 1 hour to get to the primary healthcare clinic.

The overall prevalence of food insecurity in this sample was 35.3% (Table 1). There were no statistically significant gender differences in reported household food insecurity ($p=0.989$). Prevalence of food insecurity increased with age and was highest among adults older than 37 years (45.5%, $p<0.001$). Household food insecurity was also more prevalent in adults who did not complete high school (42.4%); were currently not in school (37.1%) compared with those who had completed high school (24.0%); and individuals who were married or in a co-habiting relationship (27.3%) compared with those who were married or in a relationship but not co-habiting (22.4%).

Table 1: Description of the study sample in association with food insecurity among adults presenting for HIV counselling and testing at three primary healthcare clinics in KwaZulu-Natal, South Africa, 2010–2012

Variable	Total n (%)	Food insecurity		p-value
		Never n (%)	Ever n (%)	
Number of participants	2742 (100)	1774 (64.7)	968 (35.3)	
Gender				
Female	1447 (52.8)	936 (64.7)	511 (35.3)	0.989
Male	1295 (47.2)	838 (64.7)	457 (35.3)	
Age				
<24	1000 (36.5)	709 (70.9)	291 (29.1)	<0.001
25–28	501 (18.3)	335 (66.9)	166 (33.1)	
29–31	336 (12.3)	221 (65.8)	115 (34.2)	
32–36	349 (12.7)	206 (59.0)	143 (41.0)	
≥37	556 (20.3)	303 (54.5)	253 (45.5)	
Completed high school				
No	1684 (61.5)	970 (57.6)	714 (42.4)	<0.001
Yes	1054 (38.5)	801 (76.0)	253 (24.0)	
Currently studying				
No	2388 (87.4)	1501 (62.9)	887 (37.1)	<0.001
Yes	344 (12.6)	269 (78.2)	75 (21.8)	
Relationship status				
Not married or in a relationship	353 (12.9)	262 (74.2)	91 (25.8)	0.030
Married or in a relationship, not living together	1716 (62.6)	1131 (77.6)	385 (22.4)	
Married or in a relationship, living together	673 (24.5)	489 (72.7)	184 (27.3)	
Married				
No	2522 (92.0)	1631 (64.7)	891 (35.3)	0.922
Yes	220 (8.0)	143 (65.0)	77 (35.0)	
Employment status				
Unemployed	1785 (65.2)	1108 (62.1)	677 (37.9)	<0.001
Employed	954 (34.8)	663 (69.5)	291 (30.5)	
Do you or anyone in your household receive a government grant?				
No	1012 (36.9)	667 (65.9)	345 (34.1)	0.310
Yes	1730 (63.1)	1107 (64.0)	623 (36.0)	
Primary source of income in last 12 months?				
None	245 (8.9)	165 (67.4)	80 (32.7)	<0.001
Family	975 (35.6)	670 (68.7)	305 (31.3)	
Employment	881 (32.1)	596 (67.7)	285 (32.4)	
Spouse, partner	119 (4.3)	74 (62.2)	45 (37.8)	
Grant	481 (17.6)	247 (51.4)	234 (48.7)	
Other	40 (1.5)	21 (52.5)	19 (47.5)	
Are you currently caring financially for any dependants?				
No	1138 (41.5)	788 (69.2)	350 (30.8)	<0.001
Yes – children only	875 (31.9)	526 (60.1)	349 (39.9)	
Yes – adults only	94 (3.4)	57 (60.6)	37 (39.4)	
Yes – children and adults	634 (23.1)	403 (63.6)	231 (36.4)	
If a person in the household became ill and you needed R100 for treatment, how difficult would it be to get the money?				
Not very difficult	1493 (54.6)	1050 (70.3)	443 (29.7)	<0.001
Very difficult	1242 (45.4)	720 (58.0)	522 (42.0)	
How long did it take you to travel to the clinic today?				
Less than 1 hour	2644 (96.4)	1707 (64.6)	937 (35.4)	0.439
More than 1 hour	98 (3.6)	67 (68.4)	31 (31.6)	

Regarding socio-economic factors, experience of food insecurity was highest among unemployed participants (37.9%) and adults currently financially caring for children (39.9%). Table 1 further shows that 48.7% of individuals whose primary source of income was from government grants reported experiencing some household food insecurity compared to 32.7% among individuals with no income, or family (31.3%) or employment (32.4%) as their primary income source. Experiencing household food insecurity was similarly high (42.0%) in adults who found it very difficult to borrow ZAR100 if needed in situations of ill health.

Table 2 shows the unadjusted odds ratio (OR) of ever having been tested associated with food insecurity and the OR adjusting for confounding variables (gender, age, relationship status, employment status, primary source of income, financially supporting other adults or children, and travel time to clinic). We found that household food insecurity was not significantly associated with history of previous HIV testing (OR=0.95, 95% confidence interval (CI)=0.80–1.12, $p=0.520$).

Table 2 also presents the unadjusted and adjusted ORs for the association between household food insecurity and testing HIV positive. In both the unadjusted (OR=1.70, 95%CI=1.43–2.01) and adjusted models (OR=1.46, 95%CI=1.22–1.75), individuals who had ever experienced household food insecurity had higher odds of testing HIV positive when they presented for HCT at primary healthcare facilities.

Discussion

As UNAIDS has set ambitious global targets of 90-90-90 to be reached by 2020 in order to end AIDS as a global pandemic by 2030, achieving or exceeding the first 90% (that is, 90% of all adults knowing their HIV status) is key to ending the epidemic given the expected fallout along the cascade steps from testing to treatment and viral suppression.^{22,23} We examined the links between household food insecurity and previous HIV testing and testing HIV positive to ascertain if food insecurity is a barrier to HIV testing or a risk factor for HIV infection. We found that 35.3% of study participants came from households with some degree of food insufficiency. Although there was no association between household food insecurity and previous HIV testing, individuals who were food insecure had significantly higher odds of testing HIV positive [OR 1.47 (1.22–1.77)].

Our study revealed that 29.1% of those reporting household food insecurity were young adults (<24 years old). The incidence of HIV in South Africa is highest among young women below age 25.²⁴⁻²⁶ The majority (61.9%) of participants in our study had not completed high school. Experience of food insecurity was higher among participants who had not completed high school compared to those who had completed

high school (42.4% versus 24.0%, respectively). Weiser et al.¹⁶ found that educational level did not significantly influence the association between food insecurity and risky sexual behaviour. However, De Swardt et al.²⁷ found a positive relationship between educational level and strategies to survive food crises among South African households. Higher educational level was shown to be associated with greater ability to survive a food crisis.²⁷ Elsewhere it has further been shown that completing secondary education helps to protect against HIV acquisition, more so among young girls.^{18,28}

HIV and food insecurity are leading causes of morbidity and mortality in sub-Saharan Africa. Eradication of both HIV and food insecurity is central to the Sustainable Development Goals.²⁹ The first three of these goals are to: (1) end poverty of every form; (2) end hunger, ensure food security and better nutrition through sustainable farming; and (3) ensure health and well-being for all. Despite South Africa willingly consenting to the SDGs, the proportions of at-risk and food-insecure households have remained largely unchanged over the years.^{4,30} Considering the association between food insecurity and high-risk sexual behaviour that has been demonstrated by others among women in southern Africa¹⁶, it is a likely contributory factor to the continued high HIV risk among young women and may explain the association we found in this study^{17,20,21}.

There are some limitations to this study that should be mentioned. This analysis relied on self-reports of food sufficiency in the household by an individual member of the household. In addition, study participants were interviewed about food security cross-sectionally. Thus, we were unable to determine whether food insecurity 'caused' someone to become HIV positive or whether being HIV positive created a condition of food insecurity. A more robust design would have been to conduct a longitudinal cohort study using food diaries. This approach was, however, not feasible and beyond the scope of the Pathways to Care study. The study was also limited by the use of a single question to assess food insecurity. Finally, we did not have data that would have enabled us to explore sexual risk behaviour as a possible mediator of the association that we found between HIV positive status and food insecurity.

Despite these limitations, this analysis makes an important contribution to the literature. A strength of the analysis is that study participants were newly diagnosed HIV-positive individuals; therefore, we could show the likely role of food insecurity in seeking HIV counselling and testing. Findings of this study, together with other available literature, support the need for socio-economic and structural interventions that will transform households experiencing food insecurity into food-secure households so as to reduce household members' vulnerability to HIV acquisition. This recommendation is consistent with the Sustainable Development Goals. Failure or lack of such prudent interventions is likely to contribute to the failure to achieve the UNAIDS 90-90-90 goals.

Table 2: Association between household food security and history of previous HIV testing, and testing HIV positive

Variable	Category	Unadjusted odds ratio (95% CI)	p-value	Adjusted odds ratio (95% CI)	p-value
Effect of food insecurity on previous testing[†]					
Food insecure	Never	1		1	
	Ever	0.91 (0.77–1.06)	0.220	0.95 (0.80–1.12)	0.520
Effect of food insecurity on positive HIV status^{††}					
Food insecure	Negative	1		1	
	Positive	1.70 (1.43–2.01)	<0.001	1.46 (1.22–1.75)	<0.001

[†]Adjusted for education, employment status, currently studying, primary source of income, supports other adults or children, anyone in household receiving grants, difficulty in obtaining ZAR100 for treatment; n=2717.

^{††}Adjusted for age group, education, relationship status, employment status, currently studying, primary source of income; n=2678.

Conclusion

Household food insecurity is related to similar risk factors for HIV, such as low education, not attending school, not being married and low socio-economic status.^{25,31,32} Although we did not find an association between food insecurity and prior HCT, we found that testing HIV positive was associated with household food insecurity among adults attending HCT at public primary healthcare facilities. These findings highlight the important role that food insecurity plays in HIV risk, and are supported by findings from other sub-Saharan African countries.^{16,17} Thus, interventions such as targeted food support in poor and food-insecure populations are key to mitigating vulnerability to HIV infection. In South Africa, such an intervention could be implemented with similar means testing, as the child support or old age grants.^{33,34} If left unaddressed, household food insecurity is likely to adversely affect the global targets of ending AIDS by 2030 through the 90-90-90 and test-and-treat initiatives.

Acknowledgements

The contributions of individuals who participated in this research are greatly acknowledged. This research was supported by the US National Institute of Mental Health (R01 MH83561 and R01 MH083561-03S1; Principal Investigator: Susie Hoffman, DrPH), and by a grant from the National Institute of Mental Health to the HIV Center for Clinical and Behavioral Studies at NY State Psychiatric Institute and Columbia University [P30-MH43520; Principal Investigators: Anke A. Ehrhardt (1987-2013)/Robert H. Remien, PhD (2013–2018)]. Additional support was provided by the South African Medical Research Council to M.N. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Mental Health or the South African Medical Research Council. We thank the anonymous peer reviewers of this manuscript.

Authors' contributions

M.N.: Conceptualisation, data analysis, writing the initial draft, writing revisions. R.S.: Conceptualisation, writing the initial draft, critically reviewing the writing. T.R.: Methodology, sample analysis, critically reviewing the writing. S.H. Conceptualisation, methodology, data collection, data analysis, data curation, critically reviewing the writing, writing revisions, project leadership, funding acquisition. S.D.: Data collection, critically reviewing the writing, writing revisions, funding acquisition. K.B.: Data collection, validation, critically reviewing the writing, writing revisions. T.M.E.: Conceptualisation, critically reviewing the writing, writing revisions. E.A.K.: Data collection, critically reviewing the writing, writing revisions, project leadership. J.E.M.: Conceptualisation, critically reviewing the writing, writing revisions, project leadership. G.R.: Conceptualisation, methodology, data collection, critically reviewing the writing, writing revisions, project leadership, funding acquisition.

References

1. UNAIDS. 90-90-90: An ambitious treatment target to help end the AIDS epidemic. Geneva: Joint United Nations Programme on HIV/AIDS (UNAIDS); 2014.
2. World Health Organization (WHO). Towards universal access: Scaling up priority HIV/AIDS interventions in the health sector: Progress report April 2007. Geneva: WHO; 2007.
3. UNAIDS. UNAIDS guidance for partnerships with civil society, including people living with HIV and key populations. Geneva: UNAIDS; 2011.
4. Hongoro C, Hoosain E, Maluleke T, Reddy P, Jacobs L, Shisana O, et al. The South African National Health and Nutrition Examination Survey: SANHANES-1. Cape Town: HSRC Press; 2014.
5. Weiser SD, Young SL, Cohen CR, Kushel MB, Tsai AC, Tien PC, et al. Conceptual framework for understanding the bidirectional links between food insecurity and HIV/AIDS. *Am J Clin Nutr*. 2011;94(6):1729S–1739S. <https://doi.org/10.3945/ajcn.111.012070>
6. Pellowski JA, Kalichman SC, Cherry S, Conway-Washington C, Cherry C, Grebler T, et al. The daily relationship between aspects of food insecurity and medication adherence among people living with HIV with recent experiences of hunger. *Ann Behav Med*. 2016;50(6):844–853. <https://doi.org/10.1007/s12160-016-9812-x>

7. Boyer S, Iwuji C, Gosset A, Protopopescu C, Okesola N, Plazy M, et al. Factors associated with antiretroviral treatment initiation amongst HIV-positive individuals linked to care within a universal test and treat programme: Early findings of the ANRS 12249 TasP trial in rural South Africa. *AIDS Care*. 2016;28(suppl3):39–51. <https://doi.org/10.1080/09540121.2016.1164808>
8. Aibibula W, Cox J, Hamelin AM, Mamiya H, Klein MB, Brassard P. Food insecurity and low CD4 count among HIV-infected people: A systematic review and meta-analysis. *AIDS Care*. 2016;28(12):1577–1585. <https://doi.org/10.1080/09540121.2016.1191613>
9. Singer AW, Weiser SD, McCoy SI. Does food insecurity undermine adherence to antiretroviral therapy? A systematic review. *AIDS Behav*. 2015;19(8):1510–1526. <https://doi.org/10.1007/s10461-014-0873-1>
10. Young S, Wheeler AC, McCoy SI, Weiser SD. A review of the role of food insecurity in adherence to care and treatment among adult and pediatric populations living with HIV and AIDS. *AIDS Behav*. 2014;18(5):505–515. <https://doi.org/10.1007/s10461-013-0547-4>
11. Ivers LC, Chang Y, Gregory Jerome J, Freedberg KA. Food assistance is associated with improved body mass index, food security and attendance at clinic in an HIV program in central Haiti: A prospective observational cohort study. *AIDS Res Ther*. 2010;7(33):1–8. <https://doi.org/10.1186/1742-6405-7-33>
12. Nelson K, Cunningham W, Andersen R, Harrison G, Gelberg L. Is food insufficiency associated with health status and health care utilization among adults with diabetes? *J Gen Int Med*. 2001;16(6):404–411. <https://doi.org/10.1046/j.1525-1497.2001.016006404.x>
13. Mshana GH, Wamoyi J, Busza J, Zaba B, Chagalucha J, Kaluvya S, et al. Barriers to accessing antiretroviral therapy in Kisesa, Tanzania: A qualitative study of early rural referrals to the national program. *AIDS Patient Care STDs*. 2006;20(9):649–657. <https://doi.org/10.1089/apc.2006.20.649>
14. Katz IT, Dietrich J, Tshabalala G, Essien T, Rough K, Wright AA, et al. Understanding treatment refusal among adults presenting for HIV-testing in Soweto, South Africa: A qualitative study. *AIDS Behav*. 2015;19(4):704–714. <https://doi.org/10.1007/s10461-014-0920-y>
15. Unge C, Johansson A, Zachariah R, Some D, Van Engelgem I, Ekstrom AM. Reasons for unsatisfactory acceptance of antiretroviral treatment in the urban Kibera slum, Kenya. *AIDS Care*. 2008;20(2):146–149. <https://doi.org/10.1080/09540120701513677>
16. Weiser SD, Leiter K, Bangsberg DR, Butler LM, Percy-de Korte F, Hlanze Z, et al. Food insufficiency is associated with high-risk sexual behavior among women in Botswana and Swaziland. *PLoS Med*. 2007;4(10), e260, 10 pages. <https://doi.org/10.1371/journal.pmed.0040260>
17. Tsai AC, Hung KJ, Weiser SD. Is food insecurity associated with HIV risk? Cross-sectional evidence from sexually active women in Brazil. *PLoS Med*. 2012;9(4), e1001203, 10 pages. <https://doi.org/10.1371/journal.pmed.1001203>
18. Ramjee G, Daniels B. Women and HIV in sub-Saharan Africa. *AIDS Res Ther*. 2013;10(1), Art. #30, 9 pages. <https://doi.org/10.1186/1742-6405-10-30>
19. Hoffman S, Exner TM, Lince-Deroche N, Leu CS, Phillip JL, Kelvin EA, et al. Immediate blood draw for CD4+ cell count is associated with linkage to care in Durban, South Africa: Findings from Pathways to Engagement in HIV Care. *PLoS One*. 2016;11(10), e0162085, 16 pages. <https://doi.org/10.1371/journal.pone.0162085>
20. Miller CL, Bangsberg DR, Tuller DM, Senkungu J, Kawuma A, Frongillo EA, et al. Food insecurity and sexual risk in an HIV endemic community in Uganda. *AIDS Behav*. 2011;15(7):1512–1519. <https://doi.org/10.1007/s10461-010-9693-0>
21. Pascoe SJ, Langhaug LF, Mavhu W, Hargreaves J, Jaffar S, Hayes R, et al. Poverty, food insufficiency and HIV infection and sexual behaviour among young rural Zimbabwean women. *PLoS One*. 2015;10(1), e0115290, 21 pages. <https://doi.org/10.1371/journal.pone.0115290>
22. MacPherson P, Corbett EL, Makombe SD, Van Oosterhout JJ, Manda E, Choko AT, et al. Determinants and consequences of failure of linkage to antiretroviral therapy at primary care level in Blantyre, Malawi: A prospective cohort study. *PLoS One*. 2012;7(9), e44794, 9 pages. <https://doi.org/10.1371/journal.pone.0044794>
23. Plazy M, Dray-Spira R, Orne-Gliemann J, Dabis F, Newell ML. Continuum in HIV care from entry to ART initiation in rural KwaZulu-Natal, South Africa. *Trop Med Int Health*. 2014;19(6):680–689. <https://doi.org/10.1111/tmi.12301>

24. Shisana O, Rehle T, Simbayi LC, Zuma K, Jooste S, Zungu N, et al. South African National HIV Prevalence, Incidence and Behaviour Survey, 2012. Cape Town: HSRC Press; 2014.
25. UNAIDS. Global AIDS update 2016. Geneva: Joint United Nations Programme on HIV/AIDS (UNAIDS); 2016.
26. Ramjee G, Moonsamy S, Abbai NS, Wand H. Individual and population level impact of key HIV risk factors on HIV incidence rates in Durban, South Africa. *PLoS One*. 2016;11(4), e0153969,11 pages. <https://doi.org/10.1371/journal.pone.0153969>
27. De Swardt C, Puoane T, Chopra M, Du Toit A. Urban poverty in Cape Town. *Environ Urban*. 2005;17(2):101–111. <https://doi.org/10.1177/095624780501700208>
28. SANAC. National Strategic Plan on HIV, STIs and TB 2012-2016. Pretoria: South African National AIDS Council; 2011.
29. United Nations. Sustainable Development Goals: 17 goals to transform our world. New York: United Nations; 2015. Available from: <http://www.un.org/sustainabledevelopment/sustainable-development-goals/>
30. Roberts B, wa Kivulu M, Davids YD. South African social attitudes: The 2nd report: Reflections on the Age of Hope. Cape Town: HSRC Press; 2010.
31. Naicker N, Kharsany AB, Werner L, Van Loggerenberg F, Misana K, Garrett N, et al. Risk factors for HIV acquisition in high risk women in a generalised epidemic setting. *AIDS Behav*. 2015;19(7):1305–1316. <https://doi.org/10.1007/s10461-015-1002-5>
32. Chimbindi NZ, McGrath N, Herbst K, San Tint K, Newell ML. Socio-demographic determinants of condom use among sexually active young adults in rural KwaZulu-Natal, South Africa. *Open AIDS J*. 2010;4:88–95. <https://doi.org/10.2174/1874613601004010088>
33. Case A, Hosegood V, Lund F. The reach and impact of Child Support Grants: evidence from KwaZulu-Natal. *Develop South Afr*. 2005;22(4):467–482. <https://doi.org/10.1080/03768350500322925>
34. Moitse S, Hosegood V. Accessing grants: The never ending story. *Children First*. 2003;49:3–5.





AUTHORS:

Siphokazi Dada¹ 
Nadine Harker Burnhams^{1,2} 
Ria Laubscher³ 
Charles Parry^{1,4} 
Bronwyn Myers^{1,5} 

AFFILIATIONS:

¹Alcohol, Tobacco and Other Drug Research Unit, South African Medical Research Council, Cape Town, South Africa

²School of Public Health, University of Cape Town, Cape Town, South Africa

³Biostatistics Unit, South African Medical Research Council, Cape Town, South Africa

⁴Department of Psychiatry, Stellenbosch University, Cape Town, South Africa

⁵Department of Psychiatry and Mental Health, University of Cape Town, Cape Town, South Africa

CORRESPONDENCE TO:

Siphokazi Dada

EMAIL:

sdada@mrc.ac.za

DATES:

Received: 17 May 2017

Revised: 16 Jan. 2018

Accepted: 16 May 2018

Published: 11 Sep. 2018

KEYWORDS:

female; alcohol and other drug treatment; poly-substance abuse; readmissions; barriers

HOW TO CITE:

Dada S, Harker Burnhams N, Laubscher R, Parry C, Myers B. Alcohol and other drug use among women seeking substance abuse treatment in the Western Cape, South Africa. *S Afr J Sci.* 2018;114(9/10), Art. #4451, 7 pages. <https://doi.org/10.17159/sajs.2018/4451>

ARTICLE INCLUDES:

- × Supplementary material
- × Data set

FUNDING:

South African Department of Health; South African Medical Research Council

© 2018. The Author(s).
Published under a Creative Commons Attribution Licence.

Alcohol and other drug use among women seeking substance abuse treatment in the Western Cape, South Africa

Despite a high prevalence of alcohol and other drug use (AOD) in South Africa, little is known about AOD use among women, including those in treatment centres, and changes in use over time. This knowledge is important for the development of gender-sensitive interventions for the prevention and treatment of AOD problems. We aimed to describe changes in patterns of AOD use among women who accessed specialist AOD treatment centres in the Western Cape Province of South Africa. Data were collected from 51 specialist AOD treatment centres participating in the South African Community Epidemiology Network on Drug Use between 2000 and 2013. A total of 74 368 treatment episodes were recorded for the period, of which 22% involved women. The proportion of women seeking treatment increased from 4% to 11% over time. The most common primary substance of abuse among this sample was alcohol, followed by methamphetamine. Multinomial logistic regression analyses showed that young coloured (mixed race) women were almost 18 times more likely than other women to report methamphetamine as their primary substance of abuse (RRR=17.80; 95% CI=13.18–24.04). More than a quarter of women reported poly-substance abuse. Women treated for heroin were significantly more likely to be white and younger than 25 years old (RRR=1.62; 95% CI=1.19–2.21). These data portray an increasing use of AOD treatment services by women; particularly for alcohol and methamphetamine use disorders. Additional investigations into the service needs of women utilising AOD treatment is warranted.

Significance:

- The study provides the prevalence of AOD use disorders among women in substance abuse treatment facilities in the Western Cape.
- The study reflects treatment admissions for poly-substance abuse and severe AOD problems among these women.
- The findings highlight the importance of women-specific treatment programmes to encourage utilisation of substance abuse treatment services.

Background

Since its transition to democracy in 1994, South Africa has experienced a dramatic increase in the use of a variety of different substances of abuse.¹ According to the South African Stress and Health Survey, a large population-based mental health survey conducted in 2002–2003, the lifetime prevalence of alcohol and other drug (AOD) use disorders among South African adults was as high as 13.3%, with past year prevalence of AOD use disorders being 5.8%.² Among these 12-month prevalent cases, approximately 30.9% had a severe AOD use disorder,² with this condition being more likely among male than female individuals. Over the last decade, several studies point to lifetime use of alcohol or drugs being higher among men than women.³ For example, the 2003/2004 South African Demographic and Health Survey reported prevalence rates for alcohol use of 70.3% for men and 39.2% for women over the age of 15.³ However, findings on the prevalence of binge drinking among men relative to women suggest that, among current drinkers, problem-drinking rates are higher among women relative to men. For example, the Survey reported that rates for binge drinking on weekends were higher among women (25%) than men (23%).³ Similarly, a community-based study on access to treatment reported more daily use of alcohol among women than among men.⁴ Drinking may be more common among men, but these findings suggest that women drinkers have more problematic patterns of use relative to men.

While alcohol is still the most commonly used substance in South Africa, other substances such as cannabis and methamphetamine are widely used among women.^{5,6} A study conducted by Myers⁴ found that levels of methamphetamine use among men and women were very similar. In fact, 28% of female substance users reported daily use of methamphetamine compared to 22% of their male counterparts.⁴ Wechsberg et al.⁷ also reported very high levels of methamphetamine use among female substance users participating in an HIV-risk reduction intervention study. While these studies, like others, have mostly found that prevalence rates are consistently higher among men than women^{3,8}, the data suggest a narrowing of this gender gap in many countries^{9,10}, particularly among young women. Research conducted in treatment centres has also shown that for certain substances of abuse, such as over-the-counter and prescription medicines, the demand for treatment is greater for female users.¹¹ Despite overall lower prevalence rates when compared to men, AOD use is important to address among women as women have shorter trajectories to developing AOD use disorders than men.¹² Women with AOD problems are also more likely to experience health problems such as liver disease¹³, osteoporosis¹⁴, fertility problems and breast cancer¹⁵ than women without these problems. Additionally, women who use AODs problematically are also less likely to practise safe sex consistently with their partners⁷ and may trade sex in exchange for AODs, placing them

at risk for acquiring HIV and other sexually transmitted infections^{7,16}. Research has suggested that women who use AODs are also more vulnerable to intimate partner violence and exposure to other forms of violence than the general population.^{17,18}

Although AOD use is an important health concern for women, there has been limited South African research on this topic, including on those women who seek and access addiction treatment centres. The last few years have seen advances in local studies investigating AOD use among women in South Africa^{7,19}; however, much of the available information on prevalence has come from once-off, cross-sectional, and intervention research studies, often conducted in specific locations. As a result, AOD use trends among women have been hard to identify. Given the harmful effects of problem AOD use for women, accurate data on patterns of AOD use among women are necessary to help guide the formation and implementation of effective substance abuse policies and interventions targeting women. Through this paper, we hope to address this gap by describing trends in the nature and extent of AOD use among women who used specialist AOD treatment centres in the Western Cape between 2000 and 2013.

Methods

The South African Community Epidemiology Network on Drug Use (SACENDU) is a network of researchers, practitioners and policymakers from six sentinel sites in South Africa who meet bi-annually to provide community-level public health surveillance information about substance use related trends, including for the Western Cape. All AOD treatment centres are requested to join the network and participation within the network is voluntary. Data were collected from 51 AOD treatment sites in the Western Cape, which participate in SACENDU, representing 80% of the available treatment sites in the province. In order to be admitted to a specialist AOD treatment centre, patients are required to meet diagnostic criteria for an AOD use disorder. A standardised one-page data collection form was completed for each patient enrolled in a participating treatment centre during a 6-month period; the form consisted of 22 forced-choice questions.

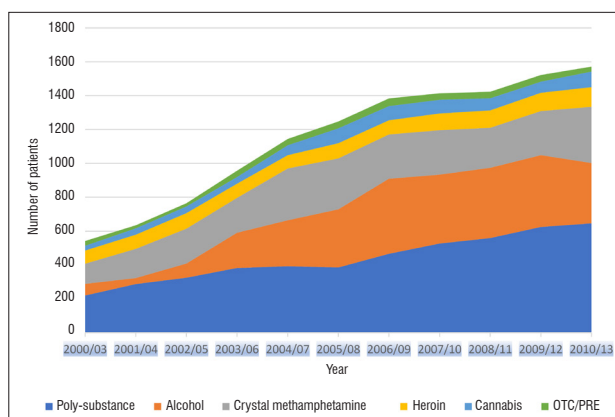
Data collection form

The data collection form contains questions about socio-demographic characteristics; referral sources; type of treatment received (inpatient or outpatient); primary and secondary substances of use (mode and frequency of use); prior treatment experiences; non-medical use of codeine-containing products; non-communicable disease diagnosis; tobacco use; HIV testing in the past 12 months, as well as service payment. Typically, the case manager completes the form a few days after the patient has been enrolled into the programme. Treatment centres are able to complete an electronic or paper form. To ensure data quality, completed forms are checked for possible miscodes and missing information before data analysis. Ethical approval for this study was provided by Stellenbosch University's Health Research Ethics Committee (N10/08/253).

Analysis

Cochran–Armitage trend tests were used to determine significant changes in female admissions to treatment between 2000 and 2013 (Figure 1). For the trend analysis, alcohol was used as a reference category. Descriptive analyses were used to describe demographic data. Chi-squared tests were performed to assess the strength of associations between participant demographic characteristics (race, marital status, education level, age category, employment status) and types of substance use (all drugs, alcohol, methamphetamine, heroin, cannabis, over-the-counter or prescription medicines and poly-substance use) of women admitted to specialist treatment centres (Table 1). A multinomial logistic regression was performed to examine the type of substance use of women with different demographic characteristics and treatment history (inpatient versus outpatient, and previous treatment). The Asian race category was excluded from this analysis because of the low number of Asians admitted to treatment centres. The variable 'type of treatment received' was used as an outcome, with 'alcohol' as the reference category. Results are presented as relative risk ratios (RRR) with 95% confidence intervals (Table 2). All statistical analyses

were conducted using Stata 13²⁰ and SAS Version 9.4²¹; $p < 0.05$ was indicative of statistical significance.



OTC/PRE, over-the-counter/prescription medicines

Figure 1: The 4-year moving average of trends in treatment demand for different (primary) substances of abuse, 2000–2013.

Results

Demographic profile and differences in patterns of substance use

From January 2000 to December 2013, 74 368 forms were collected, of which 16 656 (22.4%) were for female patients. Table 1 presents the demographic characteristics of the sample. In this data collection period, the age of female patients ranged between 10 and 85 years, with the mean age being 29.5 years (s.d. = 11.80). The majority (63.3%) of patients were coloured, followed by white patients (28.3%) and black African patients (6.3%). In relation to employment status, 54.8% of women reported that they were unemployed at the time of admission, and 24.9% reported full-time employment. Of the 16 656 female patients, 71.6% had completed high school (Grade 12), with a further 13.5% patients having had tertiary education. Of the female patients enrolled during this period, 50.5% were admitted as outpatients and were referred for treatment mostly by family or friends (22.4%) or sought treatment themselves (22.8%). A total of 26% of female patients reported having received previous treatment for an AOD problem. The majority (59.1%) of patients had never been married.

Changes in substance abuse treatment admission trends for women

The number of women admitted to AOD treatment centres increased significantly between 2000 and 2013 (Figure 1). There was a gradual increase in the proportion of admissions for alcohol-related problems from 2000 (34.1%) to 2009 (37.3%), although the proportion of alcohol-related admissions decreased and remained stable from 2011. Methamphetamine admissions saw a rapid growth from 20.4% of admissions in 2000 to 34.2% of admissions in 2004, with a significant decline in 2008 (22.2%; $z = 12.25$; $p = 0.001$), plateauing in the last 3 years. Over time, the proportion of heroin-related admissions increased significantly, from 4.8% in 2000 to 5.6% in 2013 ($z = 11.52$; $p < 0.001$). Trend analysis showed a significant increase in the proportion of cannabis admissions from less than 1.9% of admissions in 2000 to 8.3% in 2013 ($z = 5.17$; $p = 0.001$). Data suggest a significant increase in the proportion of women reporting poly-substance abuse (use of more than one substance), from 26.6% of admissions in 2000 to 42.3% of admissions in 2013 ($z = -8.30$; $p < 0.001$).

Differences in types of substance use

Multinomial logistic regression analyses revealed similarities and differences in the types of substance use among female patients of different demographic characteristics (Table 2). When compared to black African patients, coloured (RRR = 3.62; 95% CI = 2.05–4.31) and white (RRR = 3.43; 95% CI = 2.83–4.16) patients were three times more likely to be admitted for poly-substance abuse than for alcohol-related problems.

Table 1: Demographic profile and patterns of drug use among female patients receiving substance abuse treatment

	All substances (n=16 656)(%)	Poly-substances (n=6107)(%)	Alcohol (n=3746)(%)	Crystal meth (n=3390)(%)	Heroin (n=1253)(%)	Cannabis (n=921)(%)	OTC/PRE (n=473)(%)	Other drugs (n=766) (%)
Race								
African	1054(6.33)	351(5.75)	465(12.41)	61(1.80)	16(1.28)	110(11.94)	8(1.69)	43(5.61)
Asian	110(0.66)	38(0.62)	16(0.43)	18(0.53)	15(1.20)	8(0.87)	7(1.48)	8(1.04)
Coloured	10 542(63.29)*	3827(62.67)	1765(46.05)	3008(88.73)	824(65.76)	650(70.58)	132(27.91)	376(49.09)
White	4709(28.27)	1799(29.46)	1492(39.83)	287(8.47)	381(30.41)	1135(4.66)	308(65.12)	307(40.08)
Not completed	241(1.44)	92(1.51)	48(1.28)	16(0.47)	17(1.36)	18(1.95)	18(3.81)	32(4.18)
Marital status								
Married	3299(19.81)	1162(19.03)	1192(31.82)	391(11.53)	171(13.65)	26(2.82)	218(46.09)	139(18.15)
Never married	9843(59.10)*	3619(59.26)	1340(35.77)	2610(76.99)	883(70.47)	833(90.45)	109(23.04)	449(58.62)
Other	3514(21.10)	1326(21.71)	1214(32.41)	389(11.47)	199(15.89)	62(6.73)	146(30.87)	178(23.23)
Divorced	-	9.53	14.98	4.42	5.11	1.95	17.76	10.97
Living together	-	8.45	9.48	5.87	9.26	2.39	5.92	10.44
Other	-	0.65	1.47	0.15	0.16	1.30	0.00	0.78
Unknown	-	0.49	0.53	0.50	0.88	0.33	1.06	0.65
Widowed	-	2.59	5.95	0.53	0.48	0.76	6.13	0.39
Education level								
None	77(0.46)	20(0.33)	39(1.04)	7(0.21)	3(0.24)	6(0.65)	1(0.21)	1(0.13)
Grades 1–7	2069(12.42)	748(12.25)	631(16.84)	280(8.26)	65(5.19)	210(22.80)	27(5.71)	108(14.10)
Grades 8–12	11 917(71.55)*	4361(71.41)	2304(61.51)	2879(84.93)	989(78.93)	634(68.84)	258(54.55)	492(64.23)
Tertiary	2243(13.47)	835(13.67)	720(19.22)	170(5.01)	140(11.17)	56(6.08)	173(36.58)	149(19.45)
Not completed	350(2.10)	143(2.34)	52(1.39)	54(1.59)	56(4.47)	15(1.63)	14(2.96)	16(2.09)
Employment status								
Working	4161(24.98)	1465(23.99)	1491(39.80)	515(15.19)	167(13.33)	90(9.77)	189(39.96)	244(31.85)
Not working	9123(54.77)*	3583(58.67)	1520(40.58)	2295(67.70)	956(76.30)	206(22.37)	174(36.79)	389(50.78)
Student	2204(13.23)	687(11.25)	169(4.51)	523(15.43)	107(8.54)	607(65.91)	6(1.27)	105(13.71)
Other	1168(7.01)	72(6.09)	566(15.11)	57(1.68)	23(1.84)	18(1.95)	104(21.99)	28(3.66)
Age group								
≤18	2932(17.60)	948(15.52)	180(4.81)	872(25.72)	170(13.57)	628(68.19)	7(1.48)	127(16.58)
19–25	4644(27.88)*	1843(30.18)	308(8.22)	1405(41.45)	672(53.63)	146(15.85)	33(6.98)	237(30.94)
26–35	4474(26.86)	1758(28.79)	1003(26.78)	928(27.37)	327(26.10)	91(9.88)	98(20.72)	269(35.12)
36+	4539(27.25)	1531(25.07)	2244(59.90)	170(5.01)	79(6.30)	52(5.65)	335(70.82)	128(16.71)
Not completed	67(0.40)	27(0.44)	11(0.29)	15(0.44)	5(0.40)	4(0.43)	0(0.00)	5(0.65)
Treatment type								
Inpatient	8239(49.47)	3133(51.30)	2462(65.72)	1218(35.93)	456(36.39)	215(23.34)	369(78.01)	386(50.39)
Outpatient	8417(50.53)*	2974(48.70)	1284(34.28)	2172(64.07)	797(63.61)	706(76.66)	104(21.99)	380(49.61)
Previous treatment								
Yes	4434(26.62)	1929(31.59)	944(25.20)	899(26.52)	223(17.80)	154(16.72)	61(12.90)	224(29.24)
No	11 859(71.20)*	4077(66.76)	2718(72.56)	2398(70.74)	1012(80.77)	735(79.80)	398(84.14)	521(68.02)
Not completed	363(2.18)	101(1.65)	84(2.24)	93(2.74)	18(1.44)	32(3.47)	14(2.96)	21(2.74)

OTC/PRE, over-the-counter/prescription medicines

*p<0.001

Patients between the ages of 26 and 35 and those 36 years of age and older were significantly more likely to have alcohol as their primary substance of abuse (RRR=0.33; 95% CI=0.26–0.42 and RRR=0.15; 95% CI=0.12–0.20, respectively) when compared with their younger counterparts, and therefore less likely to be admitted for poly-substance abuse. Women abusing poly-substances were also more likely to be unemployed than employed (RRR=1.89; 95% CI = 1.70–2.10; Table 2) and were more likely to have had a previous treatment episode (RRR=1.36; 95% CI=1.22–1.50) than women abusing alcohol.

In relation to methamphetamine use, coloured women were nearly 18 times more likely to have methamphetamine as their primary substance of abuse (RRR=17.80; 95% CI=13.18–24.04), followed by white patients (RRR=5.24; 95% CI=3.75–7.33). Compared to patients admitted for alcohol problems, those with methamphetamine problems were twice as likely to be unemployed (RRR=2.43; 95% CI=2.10–2.81). They were also more likely to have secondary or tertiary education than those admitted for alcohol-related problems (Table 2). Those admitted with methamphetamine as their primary substance of abuse were less likely to have received inpatient treatment relative to alcohol abuse patients (RRR=0.48; 95% CI=0.42–0.55).

Women admitted for heroin treatment were significantly more likely to be white (RRR=33.37; 95% CI=18.84–59.10) when compared to patients admitted for alcohol abuse; heroin abuse patients were more likely to be between the ages of 19 and 25 or younger (RRR=1.62; 95% CI=1.19–2.21) than their alcohol using counterparts. Patients admitted for heroin abuse were four times more likely to be unemployed compared to patients admitted for alcohol abuse (RRR=4.54; 95% CI=3.70–5.59). Heroin abuse patients were also three times more likely to have a tertiary education when compared to alcohol abuse patients. Patients admitted for cannabis use were predominantly younger than 19 years of age. They were significantly less likely to be in the 19–25 years (RRR=0.38; 95% CI=0.27–0.52); 26–35 years (RRR=0.11; 95% CI=0.07–0.15) or older than 36 years of age categories (RRR=0.04; 95% CI=0.03–0.06). Furthermore, they were four times more likely to be learners (RRR=4.31; 95% CI=2.97–6.25) than persons with alcohol as their primary substance of abuse.

Patients admitted with over-the-counter and prescription medications as their primary substance of abuse, were eight times more likely to be white (RRR=8.77; 95% CI=4.02–19.11) and have a tertiary education (RRR=3.40; 95% CI=2.10–5.51), when compared to patients admitted for alcohol-related problems. These patients were also less likely to have had a previous treatment admission (RRR=0.41; 95% CI=0.30–0.55).

Discussion

Our study is among the first to explore AOD trends among women in treatment for an AOD use disorder in the Western Cape. Our findings suggest that while fewer women relative to men access specialist substance abuse treatment facilities, over time there has been a gradual increase in the proportion of women utilising these services in the Western Cape. This finding is encouraging as prior studies have documented that women with substance use disorders experience many barriers to accessing treatment, including stigma, geographical, financial, childcare and other barriers.^{22–24} Additionally, it is also thought that available treatment services are not always appropriate for the unique needs of women who use substances.²⁵ This gradual increase in women's use of treatment services is likely as a result of multiple factors, including increasing rates of substance use among women in Western Cape communities, efforts to help women overcome barriers to initiating substance abuse treatment,^{23,26} and efforts to expand the availability of services for women, including the introduction of several women-specific treatment programmes.

While our findings suggest that women seek treatment for both alcohol- and drug-related problems, it is evident that alcohol is the most common substance of abuse among women who access treatment. Findings indicate that the proportion of female admissions where alcohol was the primary substance of abuse increased significantly over the last decade. An increase in the utilisation of AOD treatment services by

people with alcohol-related problems could be attributed to alcohol-related awareness campaigns run in primary health clinics and within communities that may have increased women's awareness of when alcohol use is problematic and how to access services. Additionally, it is also not unlikely that liquor industry action, such as improved marketing initiatives, has increased women's use of alcohol which may have led to greater need for and use of AOD treatment services by women. Despite the increased utilisation of treatment services by women with alcohol problems, more needs to be done to expand access to alcohol treatment for women. One option for expanding access to care would be to identify women with problematic patterns of alcohol use in community and health settings through screening, further providing women at risk for alcohol-related problems with brief interventions to increase their motivation to engage with treatment before referring them to care. Apart from programmes that seek to identify women who may benefit from treatment and link them to care, it is also imperative that treatment services consider whether their programmes are adequately meeting the needs of their female patients, particularly given that only a quarter of the treatment population are female.

Findings also point to more women accessing treatment for methamphetamine and heroin-related problems in the Western Cape. A third of the sample reported methamphetamine as their primary substance of abuse. These patients were slightly younger than patients receiving treatment for alcohol problems. This finding is not surprising as the time from first use to treatment for people with methamphetamine problems is generally shorter than that for people using other drugs. Methamphetamine abusers cannot as easily avoid seeking help from specialist treatment centres as persons who abuse alcohol, as methamphetamine is associated with major physical harm, cardiovascular complications and disability in mental health functioning, leading to psychoses, anxiety disorders and suicidal ideation.^{27,28} Since 2004, the Western Cape in particular has seen a surge in methamphetamine admissions to specialist treatment centres. Issues like gangsterism and organised crime, increased marketing, availability and low price have been associated with high levels of methamphetamine use in this region.^{29,30} This surge in use remains a concern, particularly as people abusing this drug are significantly more likely to report severe health, family and financial problems compared to people who abuse other drugs.^{28,30,31} For instance, women with methamphetamine use disorders have more symptoms of poor mental health relative to women who use other drugs²³ and women who use methamphetamine are more likely to have been victims of sexual abuse and intimate partner violence.^{17,18} If the numerous physical and mental health needs of women who use methamphetamine are not addressed within the context of AOD treatment, it is unlikely that women will view these services as appropriate, acceptable or effective. Given findings from prior research^{25,32}, these negative beliefs about the appropriateness and effectiveness of treatment may hamper women's use of AOD treatment services²³.

Admissions for heroin abuse among women were lower than for some of the other substances but have gradually increased over the 13-year period. Similar to findings from other studies³³, women using heroin were on average younger and white. Given the high number of heroin re-admissions when compared to other substances of abuse, and multiple harms associated with heroin use (both to the individual and broader society), some consideration should be given to whether the current available treatment regimens for heroin problems are effective. Furthermore, harm-reduction interventions such as medically assisted treatment should be made available for heroin users in long-term treatment facilities, including public healthcare facilities.

Poly-substance abuse in this study was also common among patients. Our findings suggest that women who use multiple substances had more occupational problems and more severe AOD problems, as reflected in their multiple treatment episodes. This is in line with studies from other contexts, which suggest that the use of multiple illegal substances increases the likelihood of the emergence of severe psychological problems, and impacts on a plethora of social concerns both directly and indirectly.

Table 2: Multinomial logistic regression assessing the association between demographic variables and different primary substances of abuse among women in treatment

Covariates	Substance of abuse*														
	Poly-substances			Crystal meth			Heroin			Cannabis			OTC/PRE		
	RRR	95% CI	p-value	RRR	95% CI	p-value	RRR	95% CI	p-value	RRR	95% CI	p-value	RRR	95% CI	p-value
Race															
African	1.00			1.00			1.00			1.00			1.00		
Coloured	3.62	2.05 – 4.31	<0.001	17.80	13.18 – 24.04	<0.001	20.51	11.79 – 35.67	<0.001	3.29	2.48 – 4.37	<0.001	4.74	2.18 – 10.30	<0.001
White	3.43	2.83 – 4.16	<0.001	5.24	3.75 – 7.33	<0.001	33.37	18.84 – 59.10	<0.001	3.20	2.24 – 4.57	<0.001	8.77	4.02 – 19.11	<0.001
Marital status															
Ever married	1.00			1.00			1.00			1.00			1.00		
Never married	1.17	1.03 – 1.33	0.013	1.46	1.23 – 1.72	<0.001	1.15	0.92 – 1.43	0.224	3.25	2.08 – 5.07	<0.001	0.68	0.50 – 0.91	0.009
Other	1.06	0.94 – 1.20	0.332	1.00	0.82 – 1.21	0.985	0.98	0.77 – 1.27	0.904	1.83	1.12 – 2.97	0.015	0.64	0.50 – 0.82	<0.001
Age group															
≤18	1.00			1.00			1.00			1.00			1.00		
19–25	1.00	0.78 – 1.28	0.991	0.86	0.66 – 1.11	0.233	1.62	1.19 – 2.21	0.002	0.38	0.27 – 0.52	<0.001	2.58	0.82 – 8.07	0.104
26–35	0.33	0.26 – 0.42	<0.001	0.23	0.18 – 0.30	<0.001	0.29	0.21 – 0.41	<0.001	0.11	0.07 – 0.15	<0.001	2.18	0.71 – 6.77	0.176
36+	0.15	0.12 – 0.20	<0.001	0.04	0.03 – 0.05	<0.001	0.04	0.03 – 0.07	<0.001	0.04	0.03 – 0.06	<0.001	2.35	0.76 – 7.31	0.140
Employment															
Employed	1.00			1.00			1.00			1.00			1.00		
Unemployed	1.89	1.70 – 2.10	<0.001	2.43	2.10 – 2.81	<0.001	4.54	3.70 – 5.59	<0.001	1.09	0.82 – 1.45	0.567	1.48	1.16 – 1.88	0.002
Student	1.24	0.96 – 1.60	0.103	1.27	0.95 – 1.69	0.102	1.48	1.02 – 2.15	0.037	4.31	2.97 – 6.25	<0.001	0.63	0.23 – 1.75	0.374
Other work	0.98	0.84 – 1.16	0.850	0.66	0.47 – 0.92	0.014	0.83	0.49 – 1.39	0.497	1.00	0.58 – 1.72	0.999	1.32	0.99 – 1.77	0.056
Education															
Primary	1.00			1.00			1.00			1.00			1.00		
Senior	1.35	1.18 – 1.55	<0.001	2.54	2.11 – 3.04	<0.001	2.91	2.18 – 3.90	<0.001	0.75	0.60 – 0.94	0.012	2.07	1.32 – 3.25	<0.001
Tertiary	1.41	1.18 – 1.68	<0.001	1.66	1.27 – 2.16	<0.001	2.89	2.02 – 4.14	<0.001	0.98	0.67 – 1.44	0.919	3.40	2.10 – 5.51	<0.001
Other	0.54	0.29 – 0.99	0.047	0.65	0.25 – 1.68	0.376	0.73	0.15 – 3.42	0.688	0.77	0.22 – 2.69	0.687	0.68	0.09 – 5.24	0.713
Previous treatment															
No	1.00			1.00			1.00			1.00			1.00		
Yes	1.36	1.22 – 1.50	<0.001	1.08	0.95 – 1.23	0.222	0.57	0.47 – 0.68	<0.001	0.78	0.63 – 0.97	0.024	0.41	0.30 – 0.55	<0.001
Treatment type															
Outpatient	1.00			1.00			1.00			1.00			1.00		
Inpatient	0.66	0.60 – 0.74	<0.001	0.48	0.42 – 0.55	<0.001	0.45	0.28 – 0.40	<0.001	0.28	0.22 – 0.34	<0.001	0.95	0.73 – 1.25	0.733
Year	0.97	0.96 – 0.98	<0.001	0.92	0.91 – 0.94	<0.001	0.90	0.88 – 0.92	<0.001	1.03	1.01 – 1.06	0.008	0.89	0.87 – 0.92	<0.001

*Reference category alcohol

OTC/PRE, over-the-counter/prescription medicines; RRR, relative risk ratio; CI, confidence interval

In addition, poly-substance use has been found to lead to intentional and unintentional injuries, death and disability as well as multiple health problems including but not limited to cardiovascular disease, diabetes and the acquisition of infectious diseases such as tuberculosis and HIV.³⁴ Given these associations, women who are admitted to treatment with poly-substance abuse problems should be carefully screened for co-occurring health and mental health difficulties and provided with access to relevant preventative and curative health services.

Our findings also show that over half of the patients reporting cannabis use as their primary substance of abuse were younger than 18 years and were students or learners. The high prevalence of cannabis-related problems among young people mirrors similar findings from South African studies of drug use among high school students.^{35,36} As early cannabis use has implications for healthy brain development from childhood to young adulthood³⁷, and could lead to poor academic functioning and school dropout, evidence-based interventions aimed at delaying and preventing the initiation of cannabis use among young women are urgently needed.

Over time, fewer women were admitted for poly-substance abuse, methamphetamine and heroin-related problems relative to alcohol use problems. One of the reasons for this finding is the stigma associated with illicit drug use by women, which is often linked to risky sexual behaviour such as trading sex in exchange for drugs and inconsistent condom use with their partners.^{7,32} Therefore, when women seek treatment, they may be more likely to report alcohol abuse compared to other drugs in fear of this perceived judgemental attitude by treatment providers.

There are several limitations to this study. First, data were only collected from specialist substance abuse treatment centres and are unlikely to represent the magnitude of AOD use disorders among women in the Western Cape. Given the limited availability of services and limited use of treatment services by women,²⁶ the number of patients recorded in this study is likely to be a significant underestimate of the actual number of women with AOD problems. In addition, data collected from these specialist treatment centres are limited to those centres that provide the data consistently. We were not able to control for double counting across centres within particular time periods and this may have slightly inflated the number of patients receiving treatment and may have biased the data towards those patients who seek AOD treatment across more than one treatment centre. Future epidemiological studies should consider adopting a more multifaceted approach to assessing the prevalence and trends in the extent of AOD use disorders among women over time, including using other more direct measures such as general household surveys as well as indirect estimation techniques.

In conclusion, our study shows that there has been a relative increase in the use of AOD treatment by women in the Western Cape over time, particularly for poly-substance and methamphetamine use disorders. Further research is needed to better understand whether treatment programmes largely designed for, and utilised mainly by, men (who typically comprise about 80% of persons in treatment) are appropriate for the increasing number of women who access these services and are perceived to be helpful for addressing women's service needs.

Acknowledgements

We wish to acknowledge the National Department of Health (Mental Health & Substance Abuse Directorate) for its longstanding financial support of SACENDU, and staff from participating treatment centres for their continued support in providing data. The research reported in this publication was also supported by the South African Medical Research Council, which supported the salaries of all the authors.

Authors' contributions

S.D. was the lead author. S.D., N.H.B., B.M. and C.P. all contributed conceptually. S.D. collected data. N.H.B. and R.L. performed analysis. All authors contributed towards the final write-up.

References

1. Parry CDH, Pithey AL. Risk behaviour and HIV among drug using population in South Africa. *Afr J Drug Alcohol Stud.* 2006;5:139–156.

2. Herman AA, Stein DJ, Seedat S, Heeringa SG, Moomal H, Williams DR. The South African Stress and Health (SASH) study: 12-month and lifetime prevalence of common mental disorders. *S Afr Med J.* 2009;99(5 Pt 2):339–344.
3. South African Department of Health, Medical Research Council, OrcMacro. South Africa Demographic and Health Survey 2003. Pretoria: Department of Health; 2007.
4. Myers B. Access to substance abuse treatment for historically underserved communities in the Cape Town Metropole. Cape Town: Medical Research Council; 2007.
5. Dada S, Burnhams N, Erasmus J, Parry C, Bhana A, Timol F, et al. Monitoring alcohol and drug abuse treatment admissions in South Africa: July–December 2015 (Phase 39). Cape Town: Medical Research Council; 2016.
6. Van Heerden MS, Grimsrud AT, Seedat S, Myer L, Williams DR, Stein DJ. Patterns of substance use in South Africa: Results from the South African Stress and Health study. *S Afr Med J.* 2009;99(5 Pt 2):358–366.
7. Wechsberg WM, Luseno WK, Karg RS, Young S, Rodman N, Myers B, et al. Alcohol, cannabis, and methamphetamine use and other risk behaviours among Black and Coloured South African women: A small randomized trial in the Western Cape. *Int J Drug Policy.* 2008;19(2):130–139. <https://doi.org/10.1016/j.drugpo.2007.11.018>
8. Stein DJ, Seedat S, Herman A, Moomal H, Heeringa SG, Kessler RC, et al. Lifetime prevalence of psychiatric disorders in South Africa. *Br J Psychiatry.* 2008;92:112–117. <https://doi.org/10.1192/bjp.bp.106.029280>
9. Wagner FA, Anthony JC. Male-female differences in the risk of progression from first use to dependence upon cannabis, cocaine, and alcohol. *Drug Alcohol Depend.* 2007;86(2–3):191–198. <https://doi.org/10.1016/j.drugalcdep.2006.06.003>
10. Wilsnack RW, Wilsnack SC, Kristjanson AF, Vogeltanz-Holm ND, Gmel G. Gender and alcohol consumption: Patterns from the multinational GENACIS project. *Addiction.* 2009;104(9):1487–500. <https://doi.org/10.1111/j.1360-0443.2009.02696.x>
11. Myers B, Siegfried N, Parry CD. Over-the-counter and prescription medicine misuse in Cape Town – Findings from specialist treatment centres. *S Afr Med J.* 2003;93(5):367–370.
12. Greenfield SF, Back SE, Lawson K, Brady KT. Substance abuse in women. *Psychiatr Clin North Am.* 2010;33(2):339–355. <https://doi.org/10.1016/j.psc.2010.01.004>
13. Myers B, Vythilingum, B. Substance use and abuse in South Africa: Women and alcohol. In: Ellis G, Stein D, Meintjies E, Thomas K, editors. Substance use and abuse in South Africa: Insights from brain and behavioural sciences. Cape Town: UCT Press; 2012. p. 71–86.
14. Sampson HW. Alcohol and other factors affecting osteoporosis risk in women. *Alcohol Res Health.* 2002;26(4):292–298.
15. Schneider M, Norman R, Parry C, Bradshaw D, Pluddemann A, Group SACRAC. Estimating the burden of disease attributable to alcohol use in South Africa in 2000. *S Afr Med J.* 2007;97(8 Pt 2):664–672.
16. Parry CD, Dewing S, Petersen P, Carney T, Needle R, Kroeger K, et al. Rapid assessment of HIV risk behavior in drug using sex workers in three cities in South Africa. *AIDS Behav.* 2009;13(5):849–859. <https://doi.org/10.1007/s10461-008-9367-3>
17. Gass JD, Stein DJ, Williams DR, Seedat S. Gender differences in risk for intimate partner violence among South African adults. *J Interpers Violence.* 2011;26(14):2764–2789. <https://doi.org/10.1177/0886260510390960>
18. Wechsberg WM, Myers B, Reed E, Carney T, Emanuel AN, Browne FA. Substance use, gender inequity, violence and sexual risk among couples in Cape Town. *Cult Health Sex.* 2013;15(10):1221–1236. <https://doi.org/10.1010/80/13691058.2013.815366>
19. Wechsberg WM, Myers B, Kline TL, Carney T, Browne FA, Novak SP. The relationship of alcohol and other drug use typologies to sex risk behaviors among vulnerable women in Cape Town, South Africa. *J AIDS Clin Res.* 2012;S1, Art. #15, 8 pages. <https://doi.org/10.4172/2155-6113.S1-015>
20. StataCorp. Stata statistical software: Release 13. College Station, TX: StataCorp; 2013.
21. SAS Institute Inc. SAS 9.4 software. Cary, NC: SAS Institute Inc; 2013.

22. Myers B, Louw J, Pasche SE. Gendered barriers to substance abuse treatment utilization among disadvantaged communities in Cape Town, South Africa. *Afr J Psychiatry*. 2011;14:146–153. <https://doi.org/10.4314/ajpsy.v14i2.7>
23. Myers B, Kline TL, Doherty IA, Carney T, Wechsberg WM. Perceived need for substance use treatment among young women from disadvantaged communities in Cape Town, South Africa. *BMC Psychiatry*. 2014;14, Art. #100, 10 pages. <https://doi.org/10.1186/1471-244X-14-100>
24. World Health Organization. Global status report on alcohol and health 2014. Geneva: WHO; 2014.
25. Reed E, Emanuael AN, Myers B, Johnson K, Wechsberg W. The relevance of social contexts and social action in reducing substance use and victimization among women participating in an HIV prevention intervention in Cape Town, South Africa. *Subst Abuse Rehabil*. 2013;4:55–64.
26. Myers BCT, Wechsberg WM. 'Not on the agenda': A qualitative study of influences on health services use among poor young women who use drugs in Cape Town, South Africa. *Int J Drug Policy*. 2016;30:52–58. <https://doi.org/10.1016/j.drugpo.2015.12.019>
27. Darke S, Kaye S, McKetin R, Duffou J. Major physical and psychological harms of methamphetamine use. *Drug Alcohol Rev*. 2008;27(3):253–262. <https://doi.org/10.1080/09595230801923702>
28. Meade CS, Towe SL, Watt MH, Lion RR, Myers B, Skinner D, et al. Addiction and treatment experiences among active methamphetamine users recruited from a township community in Cape Town, South Africa: A mixed-methods study. *Drug Alcohol Depend*. 2015;152:79–86. <https://doi.org/10.1016/j.drugalcdep.2015.04.016>
29. Hobkirk AL, Watt MH, Myers B, Skinner D, Meade CS. A qualitative study of methamphetamine initiation in Cape Town, South Africa. *Int J Drug Policy*. 2016;30:99–106. <https://doi.org/10.1016/j.drugpo.2015.10.006>
30. Kimani SM, Watt MH, Merli MG, Skinner D, Myers B, Pieterse D, et al. Respondent driven sampling is an effective method for engaging methamphetamine users in HIV prevention research in South Africa. *Drug Alcohol Depend*. 2014;143:134–140. <https://doi.org/10.1016/j.drugalcdep.2014.07.018>
31. Watt MH, Myers B, Towe SL, Meade CS. The mental health experiences and needs of methamphetamine users in Cape Town: A mixed methods study. *S Afr Med J*. 2015;105(8):685–688. <https://doi.org/10.7196/SAMJnew.7910>
32. Myers B, Fakier N, Louw J. Stigma, treatment beliefs, and substance abuse treatment use in historically disadvantaged communities. *Afr J Psychiatry*. 2009;12(3):218–222. <https://doi.org/10.4314/ajpsy.v12i3.48497>
33. Cicero TJ, Ellis MS, Surratt HL, Kurtz SP. The changing face of heroin use in the United States: A retrospective analysis of the past 50 years. *JAMA Psychiatry*. 2014;71(7):821–826. <https://doi.org/10.1001/jamapsychiatry.2014.366>
34. Harker Burnhams N, Laubscher R, Howell S, Shaw M, Erasmus J, Townsend L. Using respondent-driven sampling (RDS) to recruit illegal poly-substance users in Cape Town, South Africa: Implications and future directions. *Subst Abuse Treat Prev Policy*. 2016;11, Art. #31, 12 pages. <https://doi.org/10.1186/s13011-016-0074-1>
35. Morojele N, Myers B, Townsend L, Lombard C, Plüddemann A, Carney T, et al. Survey on substance use, risk behaviour and mental health among Grade 8–10 learners in Western Cape Provincial Schools, 2011. Cape Town: South African Medical Research Council; 2013.
36. Reddy SP, James S, Sewpaul R, Sifunda S, Ellahebokus A, Kambaran NS, et al. Umthente Uhlaba Usamila – The 3rd South African National Youth Risk Behaviour Survey 2011. Cape Town: South African Medical Research Council; 2013.
37. Jacobus J, Tapert SF. Effects of cannabis on the adolescent brain. *Curr Pharm Des*. 2014;20(13):2186–2193. <https://doi.org/10.2174/138161281131999990426>





First report of bacterial endophytes from the leaves of *Pellaea calomelanos* in South Africa

AUTHORS:

Siphiwe G. Mahlangu¹

Mahloro H. Serepa-Dlamini¹ 

AFFILIATION:

¹Department of Biotechnology and Food Technology, University of Johannesburg, Johannesburg, South Africa

CORRESPONDENCE TO:

Mahloro Serepa-Dlamini

EMAIL:

hopes@uj.ac.za

DATES:

Received: 16 Oct. 2017

Revised: 18 Jan. 2018

Accepted: 21 May 2018

Published: 11 Sep. 2018

KEYWORDS:

Pantoea; medicinal plant; phylogenetic analysis; biolog micro-station; 16S rRNA gene

HOW TO CITE:

Mahlangu SG, Serepa-Dlamini MH. First report of bacterial endophytes from the leaves of *Pellaea calomelanos* in South Africa. S Afr J Sci. 2018;114(9/10), Art. #4235, 9 pages. <https://doi.org/10.17159/sajs.2018/4235>

ARTICLE INCLUDES:

- × Supplementary material
- × Data set

FUNDING:

University of Johannesburg

Bacteria have an endosymbiotic association with plants. Previous studies have identified endophytic bacteria and their importance in biocontrol and drug development. However, most medicinal plants identified have not been assayed for bacterial endophytes. In this study, we characterised and identified bacterial endophytes from surface-sterilised leaves of *Pellaea calomelanos*, a common fern in the Limpopo and Gauteng Provinces, South Africa. Using morphological data and 16S rRNA gene sequencing, we differentiated and identified six putative endophytic bacteria, with *Pantoea* as the dominant genus; the other two identified bacteria belong to genera *Arthrobacter* and *Bacillus*. Data from this study are an addition to the previously less studied phylloplane bacteria. This study is a pilot in cataloguing bacterial endophytes from *Pellaea calomelanos*.

Significance:

- This study provides the first report of six putative bacterial endophytes from *Pellaea calomelanos*.
- Our results will pave the way for exploring the antimicrobial activity of *P. calomelanos* bacterial endophytes and whole genome comparisons between plant bacterial endophytes and plant bacterial pathogens.

Introduction

Plants have a mutualistic relationship with varied endophytes.¹ Endophytes are endosymbiotic; they are often bacterial or fungal species which colonise the plants without causing harm or pathogenic infection³ and can spend part or their entire life cycle within the plant host²⁻⁴.

Most plant species are known to host one or more endophytic microorganisms.⁵ Endophytes form a symbiotic relationship with the plants by providing a biological defence mechanism for the plant host against pathogens,²⁻⁵ through the production of secondary metabolites. These metabolites halt the growth of or attack invading antagonists or lyse-infected plant cells; furthermore, the metabolites can induce plant host defence mechanisms and promote plant growth.⁶

The present study was designed to isolate, identify and characterise endophytic bacteria from *Pellaea calomelanos* obtained in South Africa. Pteridaceae is a family of fern plants with over 45 genera and more than 1000 species.⁷ One such genus is *Pellaea* with over 35 described species found growing in arid rocky regions and within narrow open pockets in the soil.^{7,8} *P. calomelanos* is a common fern species in the Limpopo and Gauteng Provinces of South Africa, but also grows throughout the rest of the country.⁹ Common names of *P. calomelanos* in South Africa include *inkomankomo* (Zulu), *lehorometso* (Sotho), *legogoana* (Tswana), *phalatjane* (Sepedi) and hard fern (English).⁹

P. calomelanos is a multipurpose medicinal plant used for the treatment of headaches, chest colds, asthma, head colds and mouth and nasal ulcers.⁹ Like other plants, *Pellaea* species have a mutualistic relationship with endophytes, although, to date, endophytes have been reported in only *P. concolor* and *P. viridis*.^{10,11} We thus report here on the occurrence of bacterial endophytes within the leaves of *P. calomelanos*. Ours is the first study to describe the isolation, identification and characterisation of bacterial endophytes from *P. calomelanos* using morphological and phenotypic characteristics and the sequencing of the 16S rRNA gene.

Materials and methods

Collection of plant material

Aerial portions of the plant material were collected from Botlokwa (23°29'34.8"S, 29°42'11.2"E) in the Limpopo Province of South Africa. Whole plants were placed in sterile polyethylene bags and transported to the laboratory under 4 °C. The plant material was collected in March 2017 from a site with sandy loam soil.

Identification of the plant material

The identification of the plant material was carried out at the University of Johannesburg Herbarium (JRAU). A sample specimen of the plant material was deposited in the University of Johannesburg Herbarium (JRAU) with voucher specimen number Serepa-Dlamini 201 and species name *Pellaea calomelanos*. The remaining collected plant material was immediately processed in the laboratory.

Isolation of endophytic bacteria

Immediately after collection of plant material in the lab, plant leaves were washed with running tap water followed by a sequential sterilisation with the following solutions: sterile distilled water for 1 min, 70% ethanol for 1.5 min, 1% sodium hypochlorite for 3 min and finally washed three times in sterile distilled water. The final washing water was then plated as control. The surface-sterilised leaves were ground in 2 mL of saline using a pestle and mortar. Under sterile conditions, the homogenate was streaked onto nutrient agar plates. The plates were incubated for 24–48 h at 28 °C and inspected daily for bacterial colony growth. Isolated colonies were re-cultured on sterile nutrient agar plates until pure colonies with uniform morphology were obtained. For each endophyte bacterial isolate, 35% glycerol (glycerol diluted in sterile distilled water) stock cultures were prepared and stored at -80 °C for future use.

Preliminary identification of endophytic bacterial isolates

Pure colonies were subjected to a Gram reaction test.¹² The Schaeffer–Fulton stain method¹³ was conducted to determine if the bacterial endophytes produced endospores. The hanging-drop method adapted from MacFaddin¹⁴ was used to determine motility of the bacterial isolates. All the prepared slides were examined using a bright-field compound light microscope (CX21FS1, Olympus Corporation, Tokyo, Japan) with 400x and 1000x magnification.

Identification of bacterial endophytes by the Biolog MicroPlate system

The following procedure was utilised to identify the putative endophyte bacterial isolates. Under sterile conditions, overnight single colonies of the isolates were sub-cultured in six separate 5 mL of 0.85% saline solution. The Biolog turbidimeter was used to monitor and measure the turbidity of the suspension until a 90–98% transmittance was reached. The suspension (150 μ L) was aliquoted into each well of the Biolog MicroPlate (with GEN III MicroPlate™; Biolog Inc., Hayward, CA, USA) and incubated at 26 °C for 24 h. The plates were scanned using the Biolog automatic system and samples identified using the Biolog software.¹⁵

Scanning electron microscopy

A scanning electron microscope was used to observe and study the shape of the endophyte bacterial isolates. Each glycerol stock of the isolated bacterial endophytes was inoculated into 5 mL nutrient broth and incubated for 48 h at 28 °C. The bacterial suspension was centrifuged at 2935 rcf for 10 min, the supernatant was discarded and bacterial cells were rinsed with sterile distilled water three times with a 5-min interval between each rinse. The bacterial cells were then fixed in 8% glutaraldehyde overnight (glutaraldehyde 25% EM grade diluted in Ringers' solution). Sterile distilled water was used to rinse the cells twice, followed by a series of dehydration with 30%, 50%, 70%, 90%, 95% and 100% ethanol at 10-min intervals and a centrifugation step at 2935 rcf for 10 min. The bacterial pellets were left to dry overnight in open Eppendorf tubes placed in a refrigerator at 4 °C. The cells were then mounted on scanning electron microscope stubs and coated with gold and viewed using the TESCAN VEGA 3 LMH (AVG9731276ZA, Warrendale, PA, USA) scanning electron microscope fitted with a digital camera.

Identification of bacterial endophytes using 16S rRNA

Extraction of genomic DNA

Genomic DNA of each bacterial isolate was extracted from nutrient agar pure colonies using a Quick-DNA™ ZR fungal/bacterial DNA MiniPrep kit (Zymo Research, Irvine, CA, USA) according to the manufacturer's protocol. The extracted DNA was quantified using the NanoDrop ND-2000 UV-Vis spectrophotometer (ThermoFisher Scientific, USA); the DNA concentration was 40 ng/ μ L prior to cleaning with ZR fungal/bacterial DNA clean and concentrator-5 (Zymo Research).

Polymerase chain reaction amplification and sequencing

The 16S rRNA gene of each bacterial isolate was amplified by polymerase chain reaction (PCR) in a 12.5- μ L reaction containing 1.5 μ L of template DNA, 1.5 μ L of each primer 5'-AGAGTTTGATCCTGGCTCAG-3' f and 5'-AAGGAGGTGATCCAAGCCGCA-3' r, 6 μ L of One Taq® 2X PCR master mix with standard buffer (20 mM Tris-HCl, 1.8 mM MgCl₂, 22 mM NH₄Cl, 22 mM KCl, 0.2 mM dNTPs, 5% glycerol, 0.06% IGEPAL® CA-630, 0.05% Tween® 20, 25 units/mL One Taq® DNA polymerase) and final volume filled up to 12.5 μ L with nuclease free water. The PCR cycle was performed using MyCycler™ Thermal Cycler (catalogue number 580BR 08389, BioRad, Hercules, CA, USA) with the following conditions: initial denaturation at 94 °C for 3 min, followed by 35 cycles of amplification: denaturation at 94 °C for 1 min, 55 °C for 1 min (annealing), 72 °C for 2 min (extension), followed by a final extension at 72 °C for 10 min.¹⁶ The PCR products were cleaned with ExoSAP-it™ (ThermoFisher Scientific) following the manufacturer's recommendations and sent for sequencing with primers to Inqaba Biotechnical Industries (Pty) Ltd, Pretoria, South Africa.

Phylogenetic analysis

The 16S rRNA gene sequences of the bacterial isolates were subjected to BLAST (v.2.6.0) analysis against the rRNA sequence database (Bacteria and Archaea) at the National Center for Biotechnology Information (NCBI) to identify the closest related bacterial species. Only bacterial species with 95–100% identity similarity were selected for phylogenetic analysis. The aligned sequences, containing the isolate and closest related bacterial species, were determined by MUSCLE¹⁷ and phylogenetic analysis carried out using the maximum likelihood method based on the Tamura–Nei model¹⁷. Positions overlapping with gaps and missing nucleotide data were eliminated. All evolutionary analyses were conducted in MEGA 7.¹⁸ *Escherichia coli* ATCC 11775T with GenBank accession number X80725 was used as the outgroup. The phylogenetic trees were reported with bootstrap percentages. The 16S rRNA gene sequences of bacterial isolates identified in the study were deposited in GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>) with the following accession numbers and names: MF613647 (*Arthrobacter* sp. strain MHSD1), MF613648 (*Pantoea* sp. strain MHSD2), MF613649 (*Bacillus infantis* strain MHSD3), MF613650 (*Pantoea* sp. strain MHSD4), MF613651 (*Pantoea ananatis* strain MHSD5) and MF613652 (*Pantoea* sp. strain MHSD6). The assigned names of the bacterial isolates were based on the BLAST homology percentages as well as phylogenetic results.

Results and discussion

Isolation and identification of endophytic bacteria

Morphological identification

The isolation and enumeration of endophytes from surface-sterilised plant material is the recommended method.^{3,4,19-23} In the current study, six different putative bacterial endophytes were isolated from the surface-sterilised leaves of *P. calomelanos* obtained from Botlokwa, Limpopo Province, South Africa. Previous studies on bacterial endophytes have focused on rhizosphere bacterial endophytes because of their major role in nutrient uptake and high diversity in soils.¹⁹⁻²¹

Preliminary morphological observations of the six isolated colonies that included the production of endospores, motility and Gram staining has enabled the grouping of the bacterial cultures into various groups, some with similar characteristics as shown in Table 1. The Gram stain results indicated that all bacteria were rod shaped and two out of the six isolates were Gram positive and the remaining four isolates were Gram negative. Only one of the Gram-positive isolates was an endospore former. All the isolates, except one Gram-positive isolate, exhibited motility. The morphological shapes were confirmed by the scanning electron microscope results (not shown) which showed uniform cells depicting pure cultures of the bacterial isolates.

Biolog MicroPlates

The phenotypic assays were performed on the isolates using the Biolog MicroPlates (GEN III MicroPlate) and the results are presented in Table 1. The bacterial isolates can be differentiated based on the utilisation of some of the carbon sources such as dextrin, maltose, sucrose, stachyose and pectin. Most of the isolates appear to hydrolyse these carbon sources, although variations were observed. Some of the results were not determined (ND), such as utilisation of N-acetyl-d-glucosamine, N-acetyl- β -d-mannosamine and N-acetyl neuraminic acid, and thus we cannot rely solely on the phenotypic tests for variation of the bacterial isolates.

The Biolog system can identify bacterial species both at genus and species levels, as well as through Gram stain reaction. The system does so by providing four top-ranked species for identification; Table 2 indicates that isolates 2, 4, 5 and 6 were Gram negative and belonged to the *Enterobacteriaceae* family. Although the species ID levels for these isolates had different species names for each (of the four top-ranked ID species), they all were identified to be of the genus *Pantoea*. The Gram reaction results obtained here correlate with those performed initially in the study. Isolates 1 and 3 were also predicted to be Gram positive; however, the family names as well as the genus and species names differed among each of the four predicted species names. This difference could be because we utilised a different growth medium from the one recommended in the MicroStation™ System/MicroLog User Guide.¹⁵

Table 1: Summary of phenotypic characteristics of bacterial endophytes isolated from *Pellaea calomelanos*

Characteristic	1	2	3	4	5	6
Gram reaction	+	-	+	-	-	-
Colony morphology	White circular	Yellow circular	Pink circular	Yellow circular	Yellow circular	Yellow circular
Endospore stain	-	-	+	-	-	-
Motility	-	+	+	+	+	+
Dextrin	+	+	+	+	+	+
Maltose	+	+	-	+	+	+
Sucrose	+	+	-	+	+	+
Stachyose	-	-	-	-	-	-
Ph 6	+	+	+	+	+	+
Ph 5	+	+	-	ND	ND	ND
D-melibiose	-	-	+	ND	ND	ND
D-salicin	+	+	-	ND	ND	ND
N-acetyl-d-glucosamine	ND	ND	-	+	+	+
N-acetyl-β-d-mannosamine	+	+	-	ND	ND	ND
N-acetyl-d-galactosamine	ND	ND	-	ND	ND	ND
N-acetyl neuramic acid	-	-	-	ND	ND	ND
1% NaCl	+	+	+	+	+	+
4% NaCl	+	+	+	+	+	+
8% NaCl	-	-	+	-	-	-
D-mannose	+	+	ND	+	+	+
D-fructose	+	+	+	+	+	+
D-galactose	+	+	ND	+	+	+
L-rhamnose	+	+	+	+	+	+
Inosine	+	+	-	+	+	+
% Sodium lactate	+	+	+	+	+	+
Glycerol	+	+	-	+	+	+
Pectin	+	+	+	+	+	+
Tween 40	-	-	-	-	-	-

Species: 1, *Arthrobacter* sp. MHSD1; 2, *Pantoea* sp. MHSD2; 3, *Bacillus infantis* strain MHSD3; 4, *Pantoea infantis* strain MHSD4; 5, *Pantoea* sp. MHSD5; 6, *Pantoea* sp. MHSD6. +, positive; - negative; ND, not determined

The MicroStation™ System/MicroLog User Guide¹⁵ states that if the top four identified species belong to the same or closely related genera, then the identification can be concluded as a positive result. The other parameters that can be considered from the identified results are: the probability (PROB), similarity (SIM) and distance (DIST). These parameters indicate the approximate degree of matching between the MicroPlate results and the corresponding database. SIM ≥ 0.5, DIST ≤ 5.0 and PROB close to 1 indicate reliability of the test results. Although the SIM and DIST results of the top ranked ID species for each isolate were greater than 0.5 and less than 5.0, respectively, the PROB for all isolates was at 0.7 or above 0.7, which is relatively close to 1, thus indicating reliable results.

Although the variation of phenotypic tests could not be concluded from the Biolog MicroPlates (GEN III MicroPlate), the system has shown to be reliable when identifying bacterial species to genus level; further identification can be supported by Gram stain reaction and colony morphology. We do, however, recommend that the use of Biolog MicroPlates (GEN III MicroPlate) be complemented with other strain identification methods, despite the current study showing the reliability of identifying bacterial species by use of the Biolog MicroPlates (GEN III MicroPlate). Furthermore, previous studies have identified and examined phenotypic characteristics of bacterial species utilising this system.²⁴⁻²⁷

Table 2: Biolog MicroPlate readings of isolated endophytic bacteria

Bacterial isolate number	Species ID	PROB	SIM	DIST	Organism type
1	<i>Arthrobacter globiformis</i>	0.771	0.612	7.145	GP-Rod
	<i>Arthrobacter oxydans</i>	0.056	0.093	8.908	GP-Rod
	<i>Arthrobacter ilicis</i>	0.034	0.047	8.04	GP-Rod
	<i>Arthrobacter cummunsii</i>	0.052	0.037	8.23	GP-Rod
2	<i>Pantoea agglomerans</i>	0.814	0.520	5.216	GN-Ent
	<i>Pantoea dispersa</i>	0.104	0.058	6.537	GN-Ent
	<i>Pantoea cyripedii</i>	0.067	0.036	6.815	GN-Ent
	<i>Pantoea eucria</i>	0.015	0.007	7.798	GN-Ent
3	<i>Bacillus cereus/thuringiensis</i>	0.562	0.562	6.418	GP-Rod
	<i>Brevibacterium linens</i>	0.160	0.160	6.720	GP-Rod
	<i>Corynebacterium xerosis</i>	0.104	0.104	7.257	GP-Rod
	<i>Micrococcus luteus</i> D	0.072	0.072	7.721	GP-Coccus
4	<i>Pantoea agglomerans</i>	0.610	0.610	5.622	GN-Ent
	<i>Pantoea dispersa</i>	0.089	0.089	6.596	GN-Ent
	<i>Pantoea cyripedii</i>	0.053	0.053	6.907	GN-Ent
	<i>Pantoea eucria</i>	0.007	0.007	8.155	GN-Ent
5	<i>Pantoea agglomerans</i>	0.575	0.575	6.260	GN-Ent
	<i>Pantoea dispersa</i>	0.102	0.102	7.504	GN-Ent
	<i>Pantoea cyripedii</i>	0.101	0.101	7.526	GN-Ent
	<i>Pantoea eucria</i>	0.047	0.047	8.486	GN-Ent
6	<i>Pantoea agglomerans</i>	0.603	0.603	5.744	GN-Ent
	<i>Pantoea dispersa</i>	0.104	0.104	7.118	GN-Ent
	<i>Pantoea cyripedii</i>	0.086	0.086	7.343	GN-Ent
	<i>Pantoea eucria</i>	0.057	0.057	7.862	GN-Ent

PROB, probability; SIM, similarity; DIS, distance between #1 and #2 species

GN, Gram negative; GP, Gram positive. The name appearing after the Gram stain result refers to either family or shape of identified bacteria; Ent, Enterobacteriaceae; Rod, rod shaped; Coccus, spherical or ovoid shaped.

Phylogenetic analysis

The 16S rRNA gene sequence lengths (in base pairs) are shown in Table 3. The NCBI BLAST search results indicated that bacterial endophyte isolate number 1 had a 100% identity similarity with *Arthrobacter* spp., bacterial endophyte isolate 3 had 100% identity similarity with *Bacillus* spp. and bacterial endophyte isolates 2, 4, 5 and 6 had 99% identity similarities with *Pantoea* spp. These results indicate that the dominating endophytic bacteria from *Pellaea calomelanos* belong to the genus *Pantoea*. The homology percentage identity was performed among all the bacterial isolates (results not shown). The lowest homology similarity percentage was observed between *Arthrobacter* vs *Bacillus*, *Arthrobacter* vs *Pantoea* and *Bacillus* vs *Pantoea* isolates, indicating that these are different species. There was, however, a 99% homology similarity between all the *Pantoea* isolates. These isolates were further resolved with phylogenetic analysis (Figures 1–6).

All the delineated phylogenetic trees for the isolates had polytomy relations with closely related species. Isolate 1 had a polytomy relationship with

undescribed *Arthrobacter* species and *Arthrobacter polychromogenes* (Figure 1) with a 47% bootstrap value. Isolate 3 had a polytomy relationship supported by a 98% bootstrap value with undescribed *Bacillus* species and other strains of *Bacillus infantis* (Figure 3). Isolates 2, 4, 5 and 6 had polytomy relationships with closely related *Pantoea* species and all showed varying bootstrap values – the phylogenetic relationships were different for each isolate indicating that these are different *Pantoea* species (Figures 2, 4, 5 and 6). A delineated tree including all the isolated *Pantoea* species showed that *Pantoea* sp. MHSD2 and *Pantoea* sp. MHSD4 had a monophyletic relation supported by a 60% bootstrap value, whereas *Pantoea* sp. MHSD5 and *Pantoea* sp. MHSD6 had a polytomy relationship with *Pantoea eucalypti* and *P. breunneri* species (Figure 7). The 16S rRNA gene sequence is the foremost molecular marker used to identify bacterial species; however, from the results obtained, it is evident that 16S rRNA does not resolve the phylogenetic and evolutionary relationships between closely related bacterial strains. The phylogeny of these closely related bacterial strains can be resolved with the use of multilocus sequence analysis.²⁸

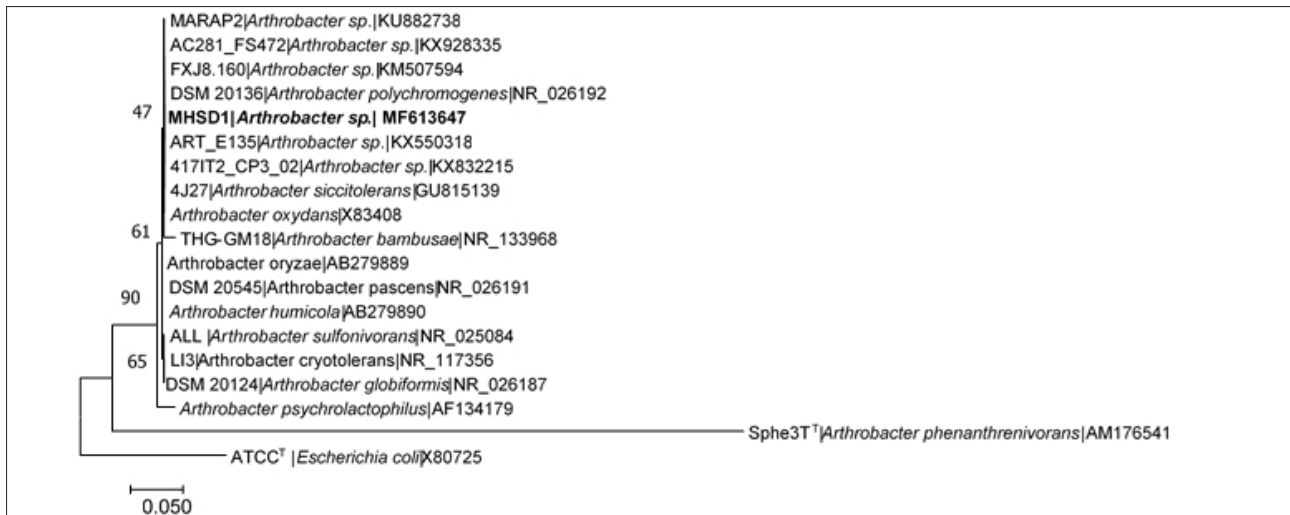


Figure 1: Maximum likelihood tree based on the 16S rRNA gene sequences of *Arthrobacter* sp. MHS1 and its closest phylogenetic neighbours. The numbers on nodes indicate bootstrap values after 1000 replicates expressed in percentages. *Escherichia coli* strain ATCC X80725 was included as an outgroup. † indicates type strains.

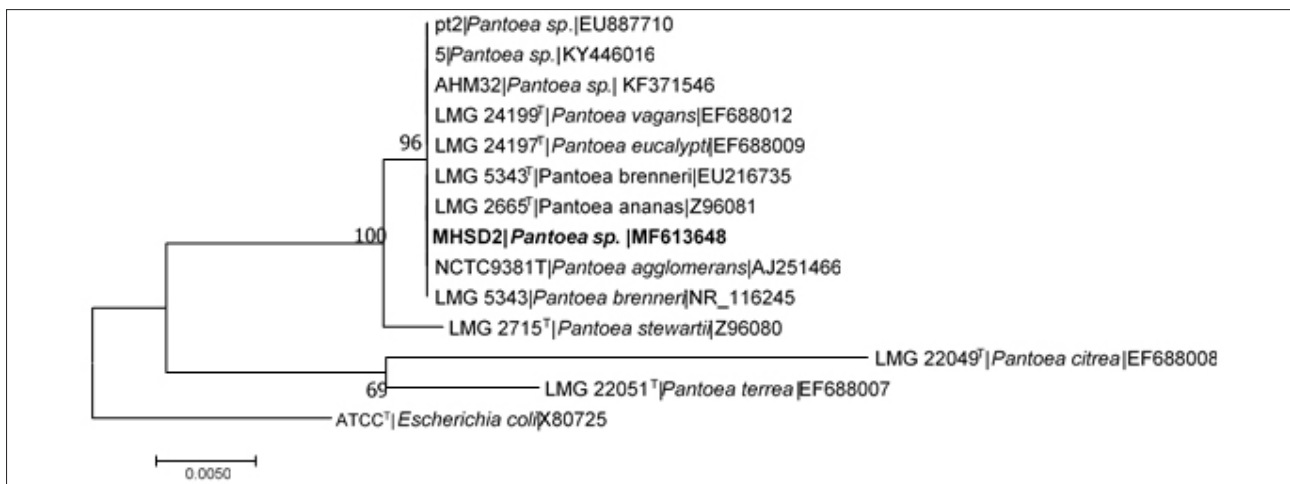


Figure 2: Maximum likelihood tree based on the 16S rRNA gene sequences of *Pantoea* sp. MHS2 and its closest phylogenetic neighbours. The numbers on nodes indicate bootstrap values after 1000 replicates expressed in percentages. *Escherichia coli* strain ATCC X80725 was included as an outgroup. † indicates type strains.

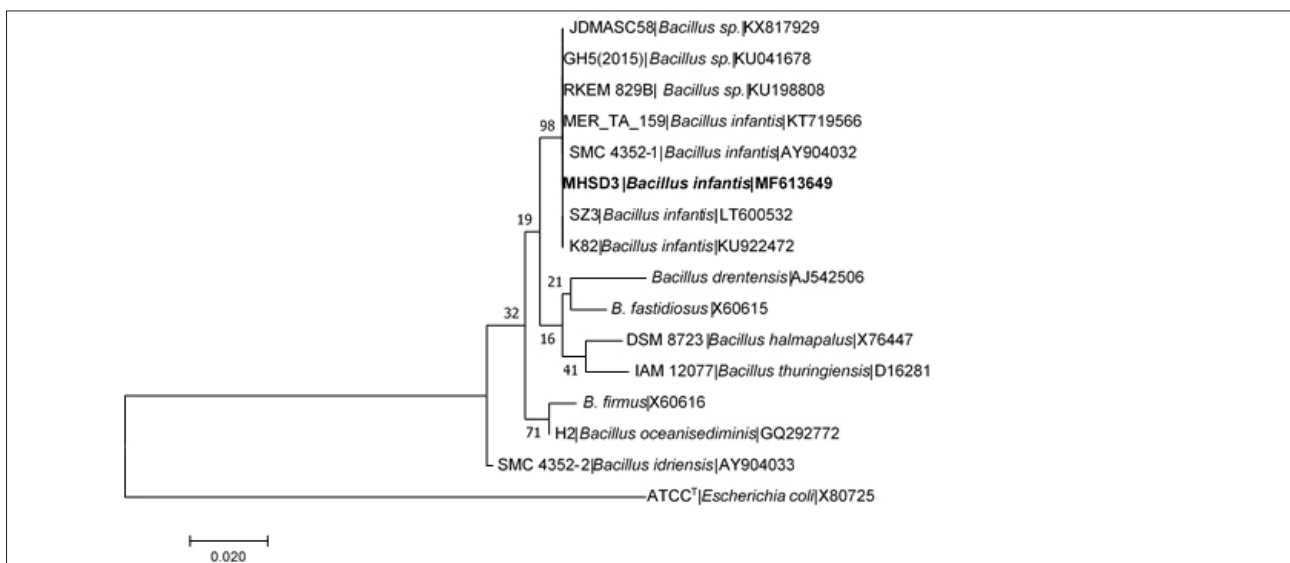


Figure 3: Maximum likelihood tree based on the 16S rRNA gene sequences of *Bacillus infantis* strain MHS3 and its closest phylogenetic neighbours. The numbers on nodes indicate bootstrap values after 1000 replicates expressed in percentages. *Escherichia coli* strain ATCC X80725 was included as an outgroup. † indicates type strains.

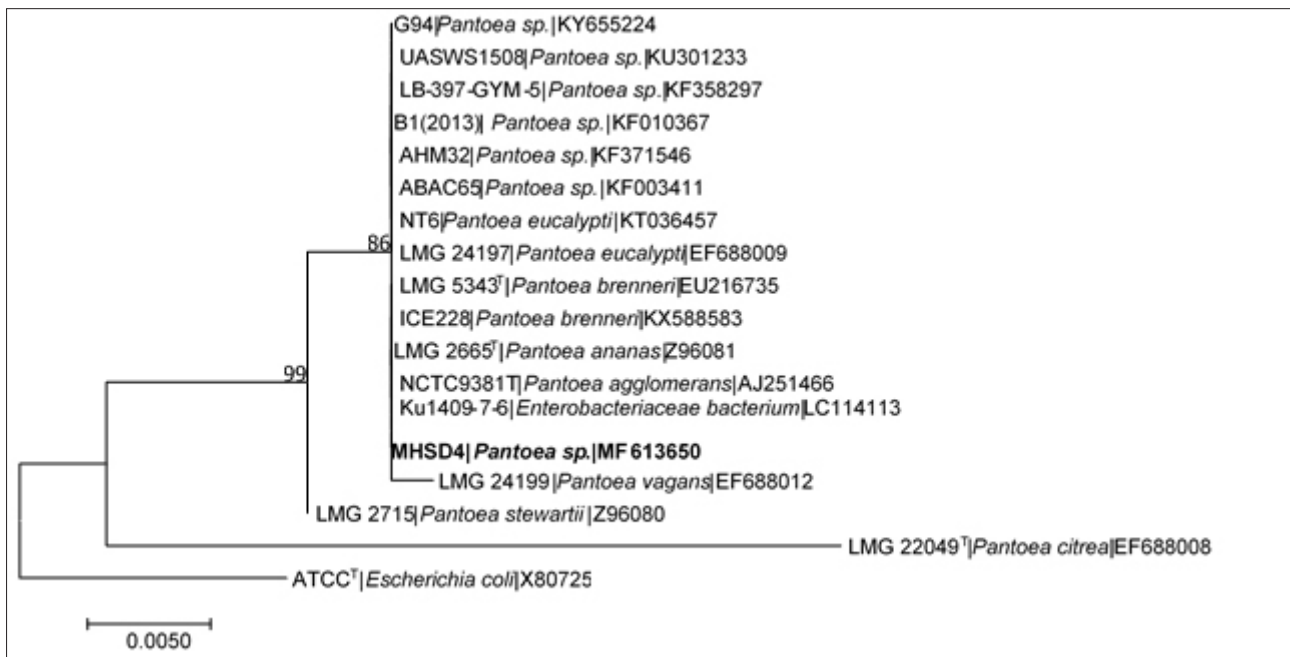


Figure 4: Maximum likelihood tree based on the 16S rRNA gene sequences of *Pantoea* sp. strain MHS4 and its closest phylogenetic neighbours. The numbers on nodes indicate bootstrap values after 1000 replicates expressed in percentages. *Escherichia coli* strain ATCC X80725 was included as an outgroup. † indicates type strains.

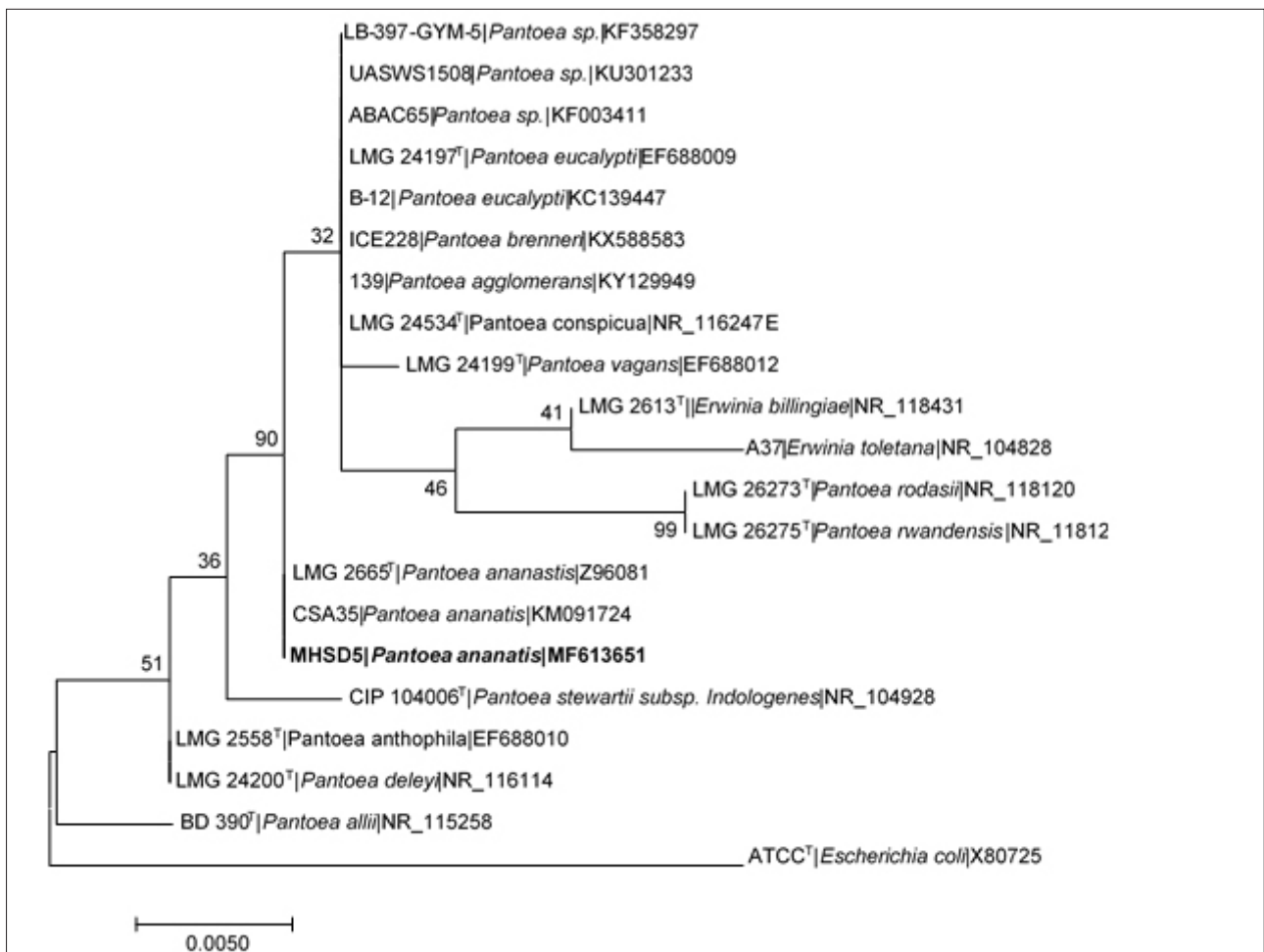


Figure 5: Maximum likelihood tree based on the 16S rRNA gene sequences of *Pantoea ananatis* strain MHS5 and its closest phylogenetic neighbours. The numbers on nodes indicate bootstrap values after 1000 replicates expressed in percentages. *Escherichia coli* strain ATCC X80725 was included as an outgroup. † indicates type strains.

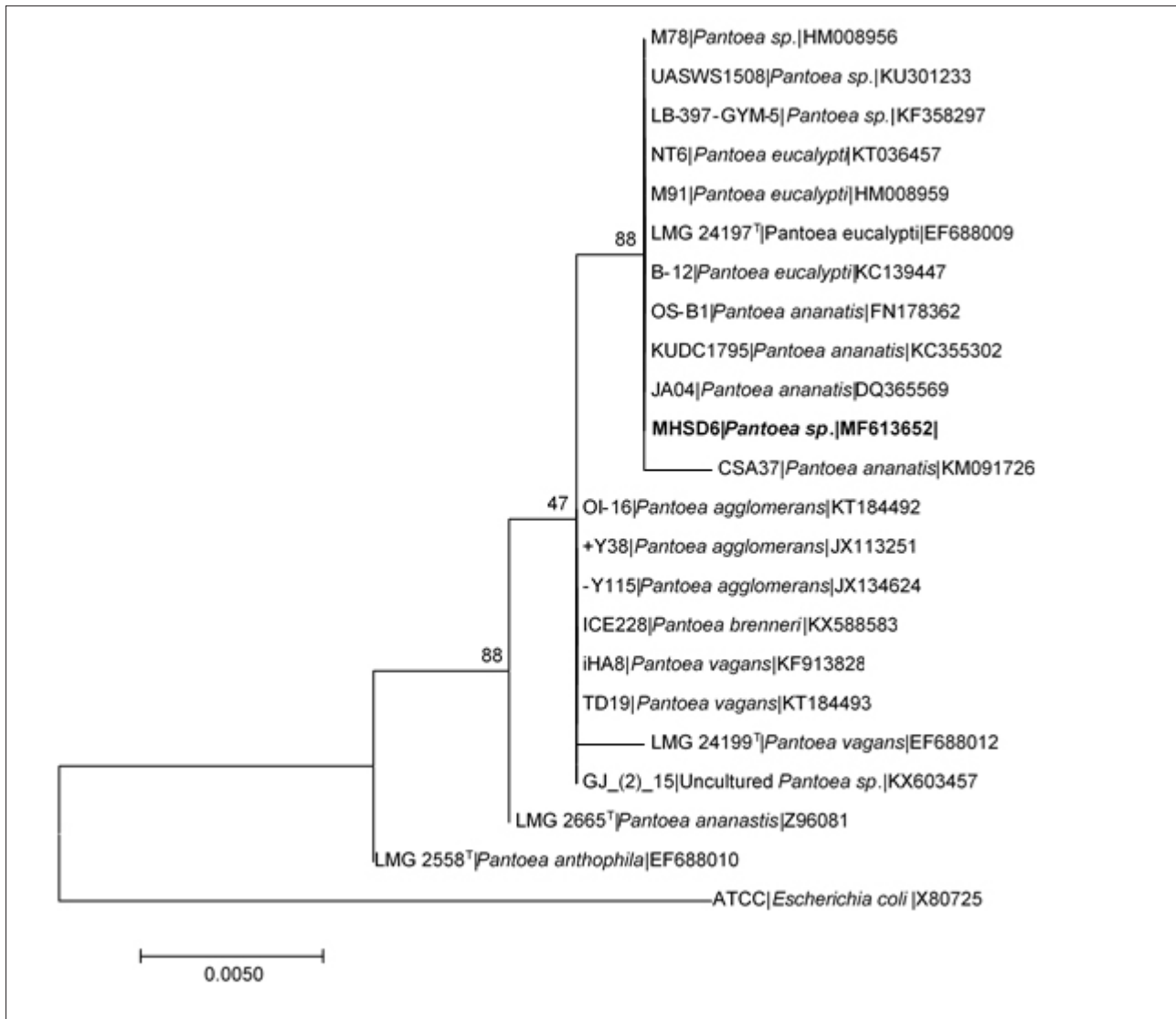


Figure 6: Maximum likelihood tree based on the 16S rRNA gene sequences of *Pantoea* sp. MHSD6 and its closest phylogenetic neighbours. The numbers on nodes indicate bootstrap values after 1000 replicates expressed in percentages. *Escherichia coli* strain ATCC X80725 was included as an outgroup. ^T indicates type strains.

Table 3: Identification of endophytic bacterial isolates

Bacterial isolate number	Size of 16S rRNA gene (base pairs)	NCBI BLAST hit results				Assigned bacterial isolate/strain name	Assigned accession number
		Dominant bacteria genus	% Query cover	E-value	% Identity		
1	629	<i>Arthrobacter</i>	100	0	100	<i>Arthrobacter</i> sp. strain MHSD1	MF613647
2	604	<i>Pantoea</i>	100	0	99	<i>Pantoea</i> sp. strain MHSD2	MF613648
3	600	<i>Bacillus</i>	100	0	100	<i>Bacillus infantis</i> strain MHSD3	MF613649
4	551	<i>Pantoea</i>	100	0	99	<i>Pantoea</i> sp. strain MHSD4	MF613650
5	557	<i>Pantoea</i>	100	0	99	<i>Pantoea ananastis</i> strain MHSD5	MF613651
6	549	<i>Pantoea</i>	100	0	99	<i>Pantoea</i> sp. strain MHSD6	MF613652

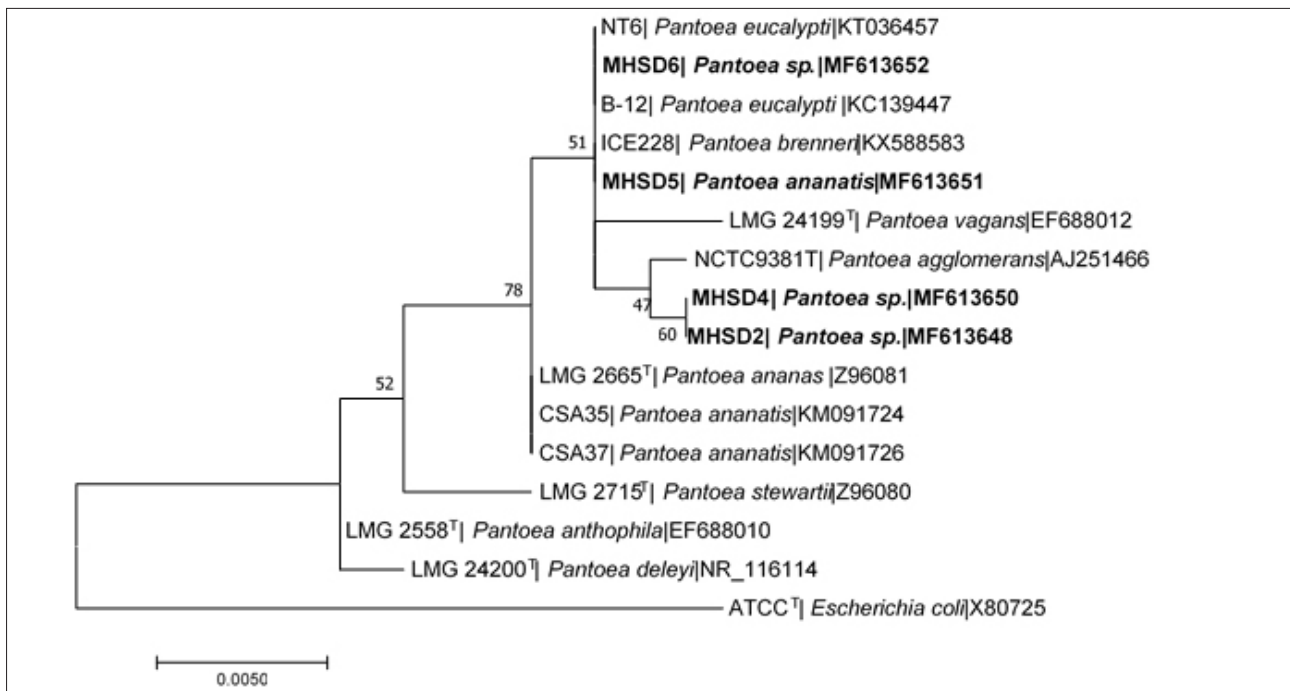


Figure 7: Maximum likelihood tree based on the 16S rRNA gene sequences of *Pantoea* spp. isolated in this study and their closest phylogenetic neighbours. The numbers on nodes indicate bootstrap values after 1000 replicates expressed in percentages. *Escherichia coli* strain ATCC X80725 was included as an outgroup. [†] indicates type strains.

We isolated bacterial endophytes from only the leaves and we believe the results obtained here will be additional to the minimally reported number of phylloplane bacterial endophytes.⁴ Braithwaite et al.²⁹ have reported on the antimicrobial activity of *P. calomelanos* against bacterial pathogens and yeast. Thus, in exploring the antimicrobial activity and other potential applications of bacterial endophytes from this plant, we had to identify and understand its associated bacterial endophytes.

Pantoea spp., *Bacillus* spp. and *Arthrobacter* spp. are common soil bacteria. Furthermore, these bacterial species have been previously isolated from maize, rice and medicinal plants as bacterial endophytes.³⁰⁻³³ Although they are considered to be plant endophytes, bacteria from the same genera as the reported bacterial endophytes have also been isolated from diseased plants and clinical samples.^{27,34} From these reports and observations, there is a need to understand the evolutionary relationships of these bacterial species, and also the occurrences and differences between bacterial endophytes as plant endophytes and as phytopathogens. Therefore, we recommend that whole genome sequencing and comparisons can provide more comprehensive insights to resolve the evolutionary relationships and identify the bacterial groups to strain level, as well as to identify genetic components that prompt the occurrence of bacterial species as bacterial endophytes.

Soil type³⁵, season^{36,37} and host type^{38,39} have been reported to affect the diversity and seasonal fluctuations of bacterial endophytes. These factors could explain the low number of bacterial endophytes isolated from the current study. Furthermore, the *Pantoea* genus was dominant, with four identified *Pantoea* species, and we strongly believe that more bacterial endophytes are likely associated with *P. calomelanos*. Further studies are underway to isolate and identify bacterial endophytes from *P. calomelanos* collected in different seasons and from different soil types. Methods which are culture independent can also be employed in identifying bacterial endophytes.

Bacterial endophytes produce the same or similar metabolites as their hosts. Therefore, because *P. calomelanos* has antibacterial and antifungal activities, the ability of bacterial endophytes from this plant to produce metabolites which (1) have therapeutic activity, (2) are

similar to those produced by *P. calomelanos* and (3) have other possible potential applications, needs to be studied.

Conclusion

Bacterial endophytes from *P. calomelanos* are poorly studied. The current study provides information on the isolation and diversity of bacterial endophytes from *P. calomelanos*. This study is a pilot to ongoing research on *P. calomelanos* obtained in South Africa, on its secondary metabolites and bacterial endophytes and potential application of the secondary metabolites and bacterial endophytes. Furthermore, whole genomic studies are underway to understand the evolutionary relationships between bacteria that occur as plant endophytes and plant pathogens. The genomic components that drive the symbiosis between *P. calomelanos* and its bacterial endophytes will also form part of future studies.

Acknowledgements

We thank the Faculty Research Committee and University Research Committee of the University of Johannesburg for funding this research. We are thankful to Mr Rudzani Manafe and Mr Witness Qaku for their technical support. Mr Sipiwe G. Mahlangu received the NRF Scarce Skills Master's scholarship (grant number SFH160628174791).

Authors' contributions

M.H.S-D. conceptualised the research; S.G.M. conducted the experiments; both authors analysed the data and contributed to the writing and editing of the manuscript.

References

- Guo LD, Hyde KD, Liew ECY. Identification of endophytic fungi from *Livistona chinensis* based on morphology and rDNA sequences. *New Phytol.* 2000;147:617-630. <https://doi.org/10.1046/j.1469-8137.2000.00716.x>
- Chojniak J, Jalowiecki L, Dorgeloh E, Hegedusova B, Ejhed H, Magner J, et al. Application of the BIOLOG system for characterization of *Serratia marcescens* ss *marcescens* isolated from onsite wastewater technology (OSWT)*. *Acta Biochim Pol.* 2015;62(4):799-805. https://doi.org/10.18388/abp.2015_1138


3. Toghueo RMK, Zabalgoeazcoa I, Vazquez de Aldana BR, Boyom FF. Enzymatic activity of endophytic fungi from medicinal plants *Terminalia catappa*, *Terminalia mantaly* and *Cananga odorata*. S Afr J Bot. 2017;109:146–153. <https://doi.org/10.1016/j.sajb.2016.12.021>
4. Gouda S, Das G, Sen SK, Shin H, Patra JK. Endophytes: A treasure house of bioactive compounds of medicinal importance. Front Microbiol. 2016;7, Art. #1538, 8 pages. <https://doi.org/10.3389/fmicb.2016.01538>
5. Strobel GA. Endophytes as sources of bioactive products. Microb Infect. 2003;5(6):535–544. [https://doi.org/10.1016/S1286-4579\(03\)00073-X](https://doi.org/10.1016/S1286-4579(03)00073-X)
6. Ding T, Melcher U. Influences of plant species, season and location of leaf endophytic bacterial communities of non-cultivated plants. PLoS ONE. 2016;11(3), e0150895, 13 pages. <https://doi.org/10.1371/journal.pone.0150895>
7. Schuetteplez E, Schneider H, Huiet L, Windham MD, Pryer KM. A molecular phylogeny of the fern family Pteridaceae: Assessing overall relationships and the affinities of previously unsampled genera. Mol Phylogenet. Evol. 2007;44(3):1172–1185. <https://doi.org/10.1016/j.ympev.2007.04.011>
8. Alan RS, Kathleen MP, Eric S, Peter K, Harald S, Paul GW. A classification for extant ferns. Taxon. 2007;55(3):705–731.
9. South African National Biodiversity Institute (SANBI). *Pellaea calomelanos* [homepage on the Internet]. c2017 [cited 2017 Jul 30]. Available from: <http://pza.sanbi.org/pellaea-calomelanos>
10. Muthukumar T, Prabha K. Abuscular mycorrhizal and septate endophyte fungal associations in lycophytes and ferns of south India. Symbiosis. 2013;59(1):15–33. <https://doi.org/10.1007/s13199-012-0185-z>
11. Tarnau K, Anielska T, Jurkiewicz A. Mycothallic/mycorrhizal symbiosis of chlorophyllous gametophytes and sporophytes of a fern, *Pellaea viridis* (Forsk.) Prantl (Pellaeaceae, Pteridales). Mycorrhiza. 2005;15(2):121–128. <https://doi.org/10.1007/s00572-004-0306-5>
12. Cruikshank R, Duguid JP, Marmoin BP, Swain RH. Medical microbiology. 12th ed. New York: Longman Group Limited; 1975.
13. Prescott LM, Harley JP, Klein DA. Laboratory exercises in microbiology. 5th ed. New York: McGraw-Hill; 2002.
14. MacFaddin J. Biochemical tests for identification of medical bacteria. 3rd ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2000.
15. MicroStation™ System/MicroLog™ user guide. A powerful automated system for the rapid identification of bacteria, yeasts and filamentous fungi. Hayward, CA; Biolog; 2010.
16. Yeates C, Gillings MR, Davison AD, Altavilla N, Veal DA. PCR amplification of crude microbial DNA extracted from soil. Lett Appl Microbiol. 1997;25:303–307. <https://doi.org/10.1046/j.1472-765X.1997.00232.x>
17. Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol. 1993;10:512–526.
18. Kumar S, Stecher G, Tamura K. MEGA 7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016;33(7):1870–1874. <https://doi.org/10.1093/molbev/msw054>
19. Jasim B, Joseph AA, John CJ, Mathew J. Isolation and characterization of plant growth promoting endophytic bacteria from the rhizome of *Zingiber officinale*. 3 Biotech. 2014;4(2):197–204. <https://doi.org/10.1007/s13205-013-0143-3>
20. Long HH, Schmidt DD, Baldwin IT. Native bacterial endophytes promote host growth in a species-specific manner; phytohormone manipulations do not result in common growth responses. PLoS One. 2008;3(7), e2702, 10 pages. <https://doi.org/10.1371/journal.pone.0002702>
21. Ashkan MF, Bleakly B. Isolation, characterization and identification of putative bacterial endophytes from some plants in Hot Springs, South Dakota. Int J Curr Microbiol App Sci. 2017;6(6):756–767. <https://doi.org/10.20546/ijcmas.2017.606.089>
22. Fouda AH, Hassan SE, Eid AM, Ewais EE. Biotechnological applications of fungal endophytes associated with medicinal plant *Asclepias sinaica* (Bioss.). Ann Agric Sci. 2015;50(1):95–104. <https://doi.org/10.1016/j.aas.2015.04.001>
23. Azevedo JL, Quecine MC. Diversity and benefits of microorganisms from the tropics. Zug, Switzerland: Springer International Publishing AG; 2017. <https://doi.org/10.1007/978-3-319-55804-2>
24. Wu W, Yang F, Bai L. Influence of substrate on electricity generation of *Shewanella loihica* PV-4 in microbial fuel cells. Microb Cel Fact. 2014;13, Art. #69, 6 pages. <https://doi.org/10.1186/1475-2859-13-69>
25. Pyar H, Peh K. Characterization and identification of *Lactobacillus acidophilus* using Biolog rapid identification system. Int J Pharm Sci. 2014;6(1):189–193.
26. Stoyanova M, Vancheva T, Moncheva P, Bogatzevska N. Differentiation of *Xanthomonas* spp. causing bacterial spot in Bulgaria based on Biolog system. Int J Microbiol. 2014;2014, Art. #495476, 7 pages. <https://doi.org/10.1155/2014/495476>
27. Brady CL, Venter SN, Cleenwerck I, Engelbeen K, Vancanneyt M, Swings J, et al. *Pantoea vagans* sp. Nov., *Pantoea eucalypti* sp. Nov., *Pantoea deleyi* sp. Nov and *Pantoea antophila* sp. Nov. Int J Syst Evol Microbiol. 2009;59:2339–2345. <https://doi.org/10.1099/ijs.0.009241-0>
28. Grevers D, Cohan FM, Lawrence JG, Spratt BG, Coenye T, Feil EJ, et al. Re-evaluating prokaryotic species. Nat Rev Microbiol. 2005;3:733–739. <https://doi.org/10.1038/nrmicro1236>
29. Braithwaite M, Van Vuuren SF, Viljoen AM. Validation of smoke inhalation therapy to treat microbial infections. J Ethnopharmacol. 2008;119(3):501–506. <https://doi.org/10.1016/j.jep.2008.07.050>
30. Xiong XQ, Liao HD, Liu XM, Zhang LY, Shi XW, Yang XL, et al. Isolation of a rice endophytic bacterium, *Pantoea* sp. Sd-1 with ligninolytic activity and characterization of its rice straw degradation ability. Lett Appl Microbiol. 2013;58:123–129. <https://doi.org/10.1111/lam.12163>
31. Shebani-Tezerji R, Naveed M, Jehl MA, Sessitsch A, Rattei T, Mitter B. The genomes of closely related *Pantoea ananatis* maize seed endophytes having different effects on the host plant differ in secretion system genes and mobile genetic elements. Front Microbiol. 2015;6, Art. #440, 16 pages. <https://doi.org/10.3389/fmicb.2015.00440>
32. Egamberdieva D, Wirth S, Behrendt U, Ahmad P, Berg G. Antimicrobial activity of medicinal plants correlates with the proportion of antagonistic endophytes. Front Microbiol. 2017;8, Art. #199, 11 pages. <https://doi.org/10.3389/fmicb.2017.00199>
33. Loh CY, Tan Y, Rohani R, Weber JF, Bhoire SJ. Diversity of endophytic bacteria in Malaysian plants as revealed by 16S rRNA encoding gene sequence based method of bacterial identification. J Young Pharm. 2013;5:95–97. <https://doi.org/10.1016/j.jyp.2013.07.001>
34. Brady C, Cleenwerck I, Venter S, Vancanneyt M, Swings J, Coutinho T. Phylogeny and identification of *Pantoea* species associated with plants, humans and the natural environment based on multilocus sequence analysis (MLSA). Syst Appl Microbiol. 2008;31:447–460. <https://doi.org/10.1016/j.syapm.2008.09.004>
35. Haridoim PR, Haridoim CC, Van Overbeek LS, Van Elsas JD. Dynamics of seed-borne rice endophytes on early plant growth stages. PLoS One. 2012;7(2), e30438, 13 pages. <https://doi.org/10.1371/journal.pone.0030438>
36. Forchetti G, Masciarelli O, Izaguirre MJ, Alemanno S, Alvarez D, Abdala G. Endophytic bacteria improve seedling growth of sunflower under water stress, produce salicylic acid, and inhibit growth of pathogenic fungi. Curr Microbiol. 2010;61:485–493. <https://doi.org/10.1007/s00284-010-9642-1>
37. Gao XX, Zhou H, Xu DY, Yu CH, Chen YQ, Qu LH. High diversity of endophytic fungi from the pharmaceutical plant, *Heterosmilax japonica* Kunth revealed by cultivation-independent approach. FEMS Microbiol Lett. 2005;249:255–266. <https://doi.org/10.1016/j.femsle.2005.06.017>
38. Ding T, Palmer MW, Melcher U. Community terminal restriction fragment length polymorphisms reveal insights into the diversity and dynamics of leaf endophytic bacteria. BMC Microbiol. 2013;13, Art. #1, 11 pages. <https://doi.org/10.1186/1471-2180-13-1>
39. Hunter PJ, Hand P, Pink D, Whipps JM, Bending GD. Both leaf properties and microbe-microbe interactions influence within-species variation in bacterial population diversity and structure in the lettuce (*Lactuca* species) phyllosphere. Appl Env Microbiol. 2010;76(24):8117–8125. <https://doi.org/10.1128/AEM.01321-10>





Mapping chlorophyll-*a* concentrations in a cyanobacteria- and algae-impacted Vaal Dam using Landsat 8 OLI data

AUTHORS:

Oupa E. Malahlela¹ 
Thando Oliphant²
Lesiba T. Tsoeleng¹
Paidamwoyo Mhangara¹

AFFILIATIONS:

¹Research and Applications Development, Earth Observation Directorate, South African National Space Agency, Pretoria, South Africa

²Data Products and Services, Earth Observation Directorate, South African National Space Agency, Pretoria, South Africa

CORRESPONDENCE TO:

Oupa Malahlela

EMAIL:

o.malahlela@gmail.com

DATES:

Received: 06 Dec. 2017

Revised: 27 Mar. 2018

Accepted: 31 May 2018

Published: 11 Sep. 2018

KEYWORDS:

chlorophyll-*a*; Landsat 8; Vaal Dam; water quality

HOW TO CITE:

Malahlela OE, Oliphant T, Tsoeleng LT, Mhangara P. Mapping chlorophyll-*a* concentrations in a cyanobacteria- and algae-impacted Vaal Dam using Landsat 8 OLI data. *S Afr J Sci.* 2018;114(9/10), Art. #4841, 9 pages. <https://doi.org/10.17159/sajs.2018/4841>

ARTICLE INCLUDES:

- × Supplementary material
- × Data set

FUNDING:

South African National Space Agency; RandWater Scientific Services

Mapping chlorophyll-*a* (chl-*a*) is crucial for water quality management in turbid and productive case II water bodies, which are largely influenced by suspended sediment and phytoplankton. Recent developments in remote sensing technology offer new avenues for water quality assessment and chl-*a* detection for inland water bodies. In this study, the red to near-infrared (NIR-red) bands were tested for the Vaal Dam in South Africa to classify chl-*a* concentrations using Landsat 8 Operational Land Imager (OLI) data for 2014–2016 by means of stepwise logistic regression (SLR). The moderate-resolution imaging spectroradiometer (MODIS) data were also used for validating chl-*a* concentration classes. The chl-*a* concentrations were classified into low and high concentrations. The SLR applied on 2014 images yielded an overall accuracy of 80% and kappa coefficient (κ) of 0.74 on April 2014 data, while an overall accuracy of 65% and $\kappa=0.30$ were obtained for the May 2015 Landsat data. There was a significant ($p<0.05$) negative correlation between chl-*a* classes and red band in all analyses, while the NIR band showed a positive correlation (0.0001; $p<0.89$) for April 2014 data set. The 2015 image classification yielded an overall accuracy of 83% and $\kappa=0.43$. The difference vegetation index showed a significant ($p<0.003$) positive correlation with chl-*a* concentrations for May 2015 and July 2016, with chl-*a* ranges of between 2.5 $\mu\text{g/L}$ and 1219 $\mu\text{g/L}$. These correlations show that a class increase in chl-*a* (from low to high) is in response to an increase in greenness within the Vaal Dam. We have demonstrated the applicability of Landsat 8 OLI data for inland water quality assessment.

Significance:

- The magnitude of the algae problem in the Vaal Dam is highlighted.
- Landsat 8 OLI satellite data have potential in mapping chl-*a* in inland water bodies.
- Both the red and the near infrared wavelengths were significant in mapping chl-*a* concentrations in the Vaal Dam.
- Satellite earth observation can be instrumental for water quality monitoring and decision-making.

Introduction

Freshwater resources are central for human sustenance and are a catalyst for economic development. However, freshwater resources globally are increasingly being polluted by industrial effluent, phosphorous and residual nitrates¹ attributed to rapid industrialisation², and the intensive use of fertilisers and pesticides in agriculture³. This situation is becoming more precarious in industrialised and agrarian economies with scarce surface freshwater resources. For many years, the subject of water quality has attracted the attention of policymakers and scientists across the African continent.^{4,5} South Africa, in particular, is currently confronted with several water quality issues. Eutrophication is at the centre of water quality challenges currently confronting water authorities in South Africa, and is a matter of public and scientific concern.⁶ Satellite-based remote sensing is increasingly playing a fundamental role in providing valuable spatial and temporal measurements of the concentration of chlorophyll-*a* (chl-*a*) in water bodies dominated by cyanobacteria and algal blooms globally. Many scientific studies have highlighted the advantages of using satellite imagery in monitoring the status of eutrophication in case II water bodies regularly.^{7,8} Thus, the estimation of chl-*a* concentration in inland water bodies is important for effective monitoring of water quality for management purposes.

However, determination of light-absorbing chl-*a* in case II water bodies (those water bodies whose optical properties are determined by phytoplankton and related constituents⁹) is more challenging considering the optical properties of the water bodies. This challenge is primarily because of the presence of other biophysical constituents such as dissolved organic matter and suspended solids with varying abundance, concentrations and reflectance.^{10,11} This variation leads to a situation in which the chl-*a* concentrations do not necessarily present a readily measurable, homogenous surface cover across the water body. Instead, complex optical properties are realised, thereby increasing challenges related to chl-*a* detection.¹¹ Fortunately, remote sensing technology has opened avenues for successful detection of chl-*a* concentrations in inland water bodies, through prevalence of algae¹² and aquatic invasive vegetation^{13–15}.

One of the most commonly used remote sensing approaches for deriving chl-*a* is hyperspectral sensing, which has proved effective for estimating chl-*a* in case II waters.^{11,16,17} However, the use of hyperspectral data and models is limited by its cost and high data dimensionality. In order to address this challenge, a number of optical multispectral sensors, such as the moderate-resolution imaging spectroradiometer (MODIS) and medium-resolution imaging spectrometer (MERIS) were used for estimating chl-*a* in inland waters.^{11,18,19} The application of MODIS and MERIS to estimate chl-*a* concentrations is possible mainly because of the relationship between chl-*a* concentrations and

spectral reflectance in red and near-infrared (NIR) regions.^{20,21} However, the coarse spatial resolution of MODIS and satellite discontinuity of MERIS have made it challenging to map chl-a to resolve small water bodies, particularly when mapping relies on the complementary nature of MODIS data by MERIS data sets. Additionally, the low signal-to-noise ratio of these sensors (8-bit) may not be adequate to characterise inland water quality.

The advancement in remote sensing technology has resulted in the launch of Landsat 5 Thematic Mapper (TM) and 7 Enhanced Thematic Mapper (ETM+) space-borne sensors with a spatial resolution of 30 m. These sensors comprise spectral regions similar to those of the MODIS sensor, in the visible to NIR region. Both Landsat 5 and 7 have demonstrated superior capability to both MERIS and MODIS in medium-size and small inland water as a result of their spatial resolution.²² They are known to be characterised by limited signal-to-noise ratio, four bands in the visible-NIR, and an 8-bit quantiser.²³ However, the instrumentation problems associated with Landsat 7 image acquisition limit the applicability of this sensor in detecting moving algal concentrations. For example, the Scan Line Corrector (SLC; an electromechanical device that compensates for the forward motion of Landsat 7) failed on 31 May 2003, resulting in up to 22% of the pixels missing in the collected image scenes.²⁴ The recently launched Landsat 8 Operational Land Imager (OLI) with improved signal-to-noise ratio and 12-bit quantisation exhibited the potential to map chl-a by using empirical band ratio regression models.²⁵ This potential is because of its stripe-free images, as opposed to the striped images acquired by Landsat 7 ETM+ over the local study area. Based on the assumptions made by Gitelson et al.²⁶ and Dall'Olmo and Gitelson²⁷ regarding the sensitivity of red and NIR bands in estimating chl-a, the question arose as to whether we can effectively use Landsat OLI data to map such chl-a concentrations in water bodies whose properties are significantly affected by mineral particles (i.e. case II water bodies).

Various statistical models have been used to map chl-a in both case I and II waters. Case I water consists of a high concentration of phytoplankton compared to other particles, whereas case II waters consist of high turbid water with an abundant occurrence of dissolved organic matter and suspended solids. Amongst these different algorithms is the linear inversion matrix²⁸, neural network methods^{13,29} and regression models³⁰. Perhaps one of the most commonly used methods is the linear regression model, which relates water/chl-a concentrations to spectral measurements as observed by satellites.³¹ There are limited studies that have used the logistic regression model, which is a type of generalised linear model, for estimation of chl-a, particularly when such estimations are done to categorise chl-a into discrete classes. Therefore, we aimed to estimate chl-a concentrations of cyanobacteria and algae in the Vaal Dam in South Africa using the Landsat 8 OLI data set. The objective of the study was to test whether single and two-band normalised indices, derived from sensitive bands, could aid in estimating chl-a concentrations in the Vaal Dam. The Vaal Dam is one of the largest and economically most important dams in South Africa, with its significance highly recognised, particularly during the dry season.³² To the best of our knowledge, there has never been a study in South Africa in which Landsat OLI data were used in modelling chl-a concentrations in large, semi-arid water bodies such as the Vaal Dam. Spatial data concerning the concentrations of chl-a in economically important dams of South Africa remain limited.

The Vaal Dam is a trans-provincial water body that forms a boundary between the Gauteng and Mpumalanga Provinces, Gauteng and Free State Provinces, and Free State and Mpumalanga Provinces of South Africa. It is located at 26°56'49"S and 28°14'30"E (Figure 1). The dam was constructed in 1938 with a wall height of 54.2 m, which was elevated to 63.5 m in 1985 because of increased water capacity of 2188 million m³/year.³³ The Vaal Dam has a surface area of approximately 321 km² and an average depth of 22.5 m; and the Dam supplies large volumes of water to the people of South Africa.³⁴ The water is mostly turbid, with the Dam characterised by silty bottom strata responsible for constant turbidity at varying degrees. Apart from the turbid nature of the Dam, algal blooms that affect the biophysical and chemical quality of the water are a growing concern³⁵; these blooms are a common

phenomenon along the Vaal River System. Because of its economic importance, there is a need to monitor algal blooms and to control cyanobacteria – which are the primary indicators of eutrophication levels in case II waters – within this reservoir. Cyanobacterial blooms have been a major environmental concern for many years, with well-known cattle deaths in the 1940s associated with cyanobacteria.³⁶ Although there is sufficient public awareness of the impacts of cyanobacteria in South Africa, information on the spatial distribution of chl-a over time within the Vaal Dam is lacking, and is crucial for reservoir management.

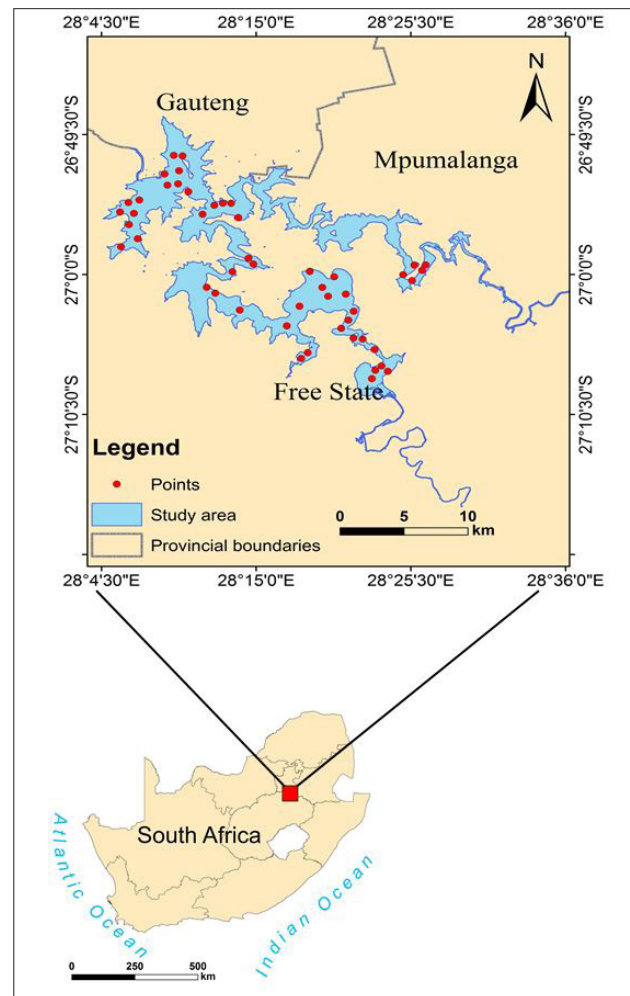


Figure 1: Locality map of the study area.

Methods

Data and pre-processing

Landsat 8 OLI data were explored as a source to estimate chl-a concentrations of cyanobacteria and algae within the Vaal Dam. Table 1 shows the details of images acquired for the study. One of the images was acquired on 22 April 2014 – the day that corresponded to one of the highest peaks of algal blooms in water bodies in the Vaal Dam.³⁷ Landsat 8 data comprise a total of 11 spectral bands in the visible to thermal infrared region (0.43–12.51 μm). In the visible shortwave infrared spectrum, Landsat 8 has a band range from coastal band (0.44 μm) to shortwave infrared (2.29 μm) and a ground sampling distance of 30 m. Each of the spectral bands in the visible shortwave infrared range was converted from a digital number to top-of-atmosphere reflectance value (0–1) by means of Equation 1:

$$\rho\lambda' = \rho_M Q_{cal} + \rho_A \quad \text{Equation 1}$$

where ρ_a is top-of-atmosphere planetary reflectance, without correction for solar angle; ρ_M is a band-specific multiplicative rescaling factor, ρ_A is the band-specific additive rescaling factor, and Q_{cal} is the quantised and calibrated standard pixel values in digital numbers.³⁸ Based on our objective, a total of seven (7) spectral bands were merged to form multispectral images with 16-bit radiometric resolution. These spectral bands were coastal (0.44 μm), blue (0.48 μm), green (0.56 μm), red (0.66 μm), near-infrared (0.87 μm), shortwave infrared1 (1.61 μm) and shortwave infrared2 (2.20 μm). The atmospheric correction processing was done on each satellite image using Quick Atmospheric Correction (QUAC) module in ENVI[®].³⁹ The resultant QUAC images had values from 0 to 10 000, and were rescaled to have values of between 0 and 1 using Equation 2:

$$lmg_n = lmg_i \times 0.0001 - 0.1 \quad \text{Equation 2}$$

where lmg_n is the new reflectance image (0–1 value) and lmg_i is the QUAC image with values scaled to 10 000. This step was necessary to allow for comparison of image values after the effect of atmosphere was removed from the images. The QUAC algorithm was chosen because of its accuracy in chl-a estimation for turbid waters and because it comes as a standard extension in the ENVI software.^{38,39}

In order to derive the geographical extent of the Vaal Dam from satellite images, a simple water index algorithm was applied on individual images. The index is based on the low within-class variation of water pixels in blue (0.48 μm) and shortwave infrared (1.61 μm) bands of Landsat 8 OLI. It classifies water pixels as 1 and non-water pixels as 0 and has shown to be effective in many parts of the world.³⁸ This processing was followed by subsetting an atmospherically corrected 7-band Landsat OLI imagery of the delineated study area. Ultimately, the pre-processing included pan-sharpening of the individual image with a 15-m panchromatic band in order to improve the ground sampling distance of the 30-m images using the nearest-neighbour resampling method.

Table 1: Specifications of the satellite data used in the study

Sensor	Path/row	Ground sampling distance (m)	Date	Atmospheric condition
OLI	170/079	15	22 April 2014	Partly cloudy
OLI	170/079	15	27 May 2015	Clear conditions
OLI	170/79	15	16 July 2016	Clear conditions
MODIS	-	250	21 April 2014	Clear conditions
MODIS	-	250	26 May 2015	Clear conditions

OLI, Operational Land Imager; MODIS, moderate-resolution imaging spectroradiometer

Selecting regions of interest

The visual interpretation of various chl-a densities was done on pan-sharpened Landsat 8 imagery (15 m ground sampling distance) in order to extract regions of interest (ROIs). The ROIs were chosen with the aid of the spectral reflectance curve shown in Figure 2 and the two-class algorithm based on the normalised difference vegetation index (NDVI) shown in Figure 3. The spectral profiles were pre-assessed using average spectra of extracted pixels from the 2016 image classes (low=17 and high=6) corresponding to field data collected during this period. Figure 2 shows that average spectra of the low and high chl-a classes for both visual (image data set) and quantitative (field data set) selection are not very different from each other. Additionally, Table 2 gives a summary of the statistics of low and high chl-a classes derived from Landsat data and field data collected on the Vaal Dam. A total of 49 ($n=49$) ROIs was extracted from each image, corresponding to images acquired in 2014 to 2016. The ROIs were chosen to fall within low or high chl-a concentration classes.

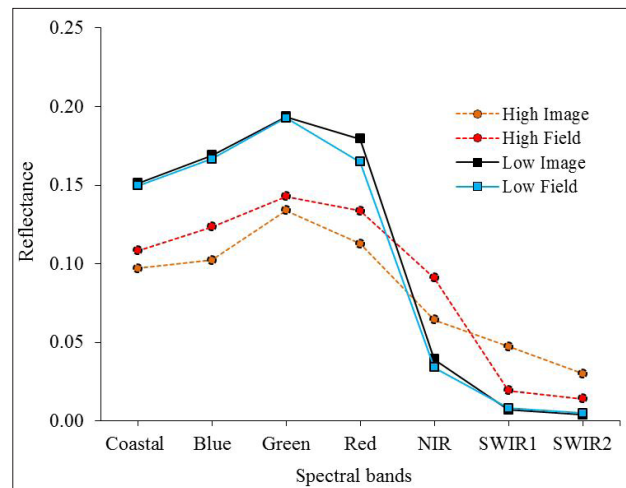


Figure 2: Typical spectral profiles of high and low classes of chl-a as observed from Landsat OLI field data of the study area.

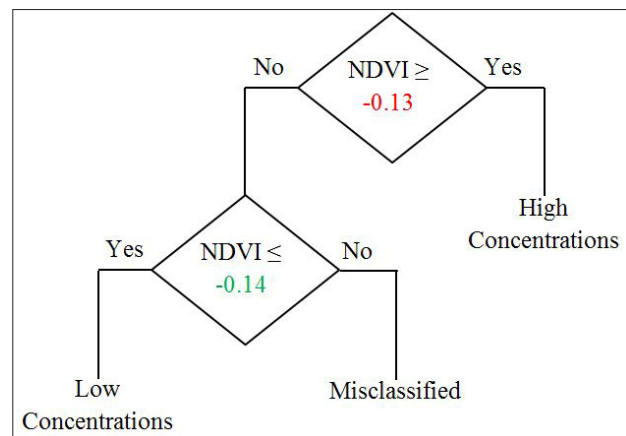


Figure 3: Normalised difference vegetation index (NDVI) classification scheme used in the study to separate chl-a concentrations into two classes.

Table 2: Descriptive statistics of two chl-a classes from the normalised difference vegetation index (NDVI) calculated using red and near-infrared bands of Landsat OLI data, and the NDVI derived from field data

	Class	Low (field)	High (field)
Descriptive statistics	Minimum	-0.23 (-0.64)	-0.13 (-0.61)
	Mean (\bar{x})	-0.22 (-0.53)	-0.04 (-0.36)
	Maximum	-0.14 (-0.45)	0.42 (-0.12)
	Standard deviation (σ)	0.018 (0.06)	0.11 (0.21)

Field data collection

Field data were collected on 16 July 2016 from the Vaal Dam. Sampling was done in terms of the exact longitudes and latitudes of the sampling points and in-situ classification of chl-a concentration; actual algae samples were taken for laboratory analysis. The data were collected corresponding to the time of the Landsat 8 OLI overpass (at approximately 10:00). A total of 23 samples was collected and all samples were analysed for chl-a concentration. Table 3 gives the details of the field data collected. This data set was used for validating the model used for predicting chl-a and was later used for producing the maps. Figure 4 shows the scatterplot of the points collected in the field. Because the

minimum chl-a value (2.5 µg/L) was far less than the maximum chl-a value (1219 µg/L), we applied the log transformation to this skewed data using $\text{Log}_x + 1$ in order to force the data to conform to normality.

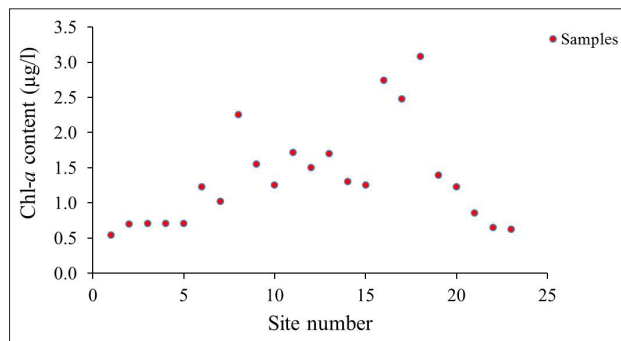


Figure 4: Scatterplot showing the distribution of chl-a samples collected from the Vaal Dam during the 2016 field campaign.

Table 3: Summary of the field data collected in July 2016

	Class	Low (µg/L)	High (µg/L)
Descriptive statistics	Minimum	2.5	49
	Mean (\bar{x})	12.7	392
	Maximum	35	1219
	Std. deviation (σ)	10.3	446.8

Data analysis

Data analysis was performed on three Landsat images in R,⁴⁰ QGIS version 2.12 and ENVI version 4.7. Various vegetation indices that are sensitive to chl-a were derived. These indices, in addition to chl-a sensitive bands, were selected on the basis of their relationship with chl-a in case II water bodies.²⁶ These indices were two-band indices in the red-NIR region of Landsat OLI. Table 4 shows the vegetation indices used in this study. From each image, 60% of the ROIs ($n=29$) was used for training the model, while 40% ($n=20$) was reserved as an independent validation data set. Similar ROIs were used for calibrating the model for the 2016 image, with the field data set serving as validation. In order to estimate classes of chl-a using remote sensing data, a multiple step-wise logistic regression (SLR) was employed. The SLR is given in the form of:

$$y_i = \frac{\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \epsilon)}{1 + \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \epsilon)} \quad \text{Equation 3}$$

where y is the resultant chl-a concentration of an i -th class, a is the y -intercept, b_n is the regression estimate of variable x_n , and ϵ is an error associated with prediction which was not pre-determined but it is associated with the logistic regression model. In addition, the D^2 (which is an analogy of R^2) was used to compare the strength of the model fit. The D^2 is given by:

$$D^2 = 1 - \left(\frac{\text{residual deviance}}{\text{null deviance}} \right) \quad \text{Equation 4}$$

Because we were using images acquired on different dates, it was necessary to assess possible differences and similarities between ROIs collected from the 2016 image (date of field data acquisition) and those collected from 2014 and 2015 images. For this reason, we adopted the spectral discrimination index (SDI) for the 2014/2016 image pair and the 2015/2016 image pair^{41,42} acquired for the same season (autumn). The SDI is computed from Equation 5 as:

$$\text{SDI} = \frac{|\mu_1 - \mu_2|}{\sigma_1 + \sigma_2} \quad \text{Equation 5}$$

where μ_1 and μ_2 are the mean values of chl-a pixels in time period 1 and 2, respectively; and σ_1 and σ_2 are the standard deviations of the chl-a pixels between time 1 and 2, respectively. A higher SDI indicates that these two images are significantly different while a lower value indicates that there is a form of similarity in conditions at times 1 and 2. A rule of thumb is that an $\text{SDI} > 1$ shows a satisfactory dissimilarity between two mean values. The preliminary SDI on extracted ROIs showed less varying chl-a spectra on Landsat 8 OLI images acquired in 2014–2016 ($\mu_{2014} = 0.090$, s.d. = 0.053; $\mu_{2015} = 0.103$, s.d. = 0.083; $\mu_{2016} = 0.084$, s.d. = 0.037). The SDI value between 2014 and 2016 images was 0.07, while a value of 0.16 was obtained on images acquired in 2015 and 2016. This information suggested that the spectral analysis could further be done for all three image dates as the conditions prevailing during field data collection did not significantly differ ($\text{SDI} < 0.2$).

Accuracy assessment

In order to validate the reliability of the results and model used in estimating two main classes of chl-a concentrations, statistical comparison was done between the predicted chl-a and visually quantified chl-a. For a more robust validation of the model, the field data collected on 16 July 2016 were used as an independent validation data set ($n=23$), which was not significantly different from the data set used for 2014 and 2015 validation ($n=20$). The NDVI derived from MODIS was used against the predicted classes. The MODIS data were used for evaluating any similarities between MODIS NDVI and Landsat 8 NDVI used to classify chl-a preliminarily into high and low classes. The NDVI is the quantitative indicator of vegetation greenness and its relative density as observed using space-borne satellites. The NDVI values have a range from -1.0 to +1.0, with positive values (close to +1.0) indicating healthy green pigment. Geographical features such as water, areas of barren rock, sand, and bare soil are usually characterised by values below zero (close to -1.0).

The MODIS NDVI product has shown correlation with various chl-a classes and the ground data in Heihe River Basin, with an R^2 of up to 0.98.⁴³ On the other hand, Gitelson et al.⁴⁴ have shown that MODIS-retrieved NDVI vs fAPAR_{green} showed a very close relationship with the fAPAR_{green} vs in-situ NDVI_{green}, while MODIS NDVI showed positive correlation with chlorophyll content⁴⁵. This makes the MODIS NDVI product a good reference data set for chl-a studies. The reference data set was obtained from points generated on MODIS NDVI with 250-m pixel resolution, and the ROIs from Landsat 8. The MODIS NDVI products were produced for the dates that correspond with the Landsat 8 image acquisitions for 2014 and 2015, as the 2016 data set could only be validated using the field data set. Because MODIS NDVI values are scaled from -1999 to 7500, we converted them to standard NDVI values of between -1.0 and 1.0 using Equation 6:

$$\text{NDVI}_n = \text{NDVI} \times 0.0001 \quad \text{Equation 6}$$

where NDVI_n is the standard NDVI, and NDVI is the MODIS NDVI product. Classification accuracy was computed from the overall accuracy, which is the total number of correctly classified cases to the total number of cases. In addition, the accuracy was further assessed by calculating the kappa coefficients (κ) between classes of low chl-a concentrations (0) and high concentrations (1) using the field chl-a data set. The κ is given by:

$$\kappa = \frac{\rho_{\text{obs}} - \rho_{\text{chn}}}{1 - \rho_{\text{chn}}} \quad \text{Equation 7}$$

where ρ_{obs} is the observed proportion of agreement and ρ_{chn} is the proportion expected by chance.

Table 4: The input remote sensing variables used for estimating chl-a

Variable (index)	Formulation	Reference	Sensitivity
β_1	$\beta_1 \alpha (R_{rs}^{-1} (0.66) \times R_{rs}^{-1} (0.86))$	Dall’Olmo and Gitelson ²⁷	Chlorophyll-a
β_2 (simple ratio)	$\beta_2 \alpha (R_{rs}^{-1} (0.86) / R_{rs}^{-1} (0.66))$	Jordan ⁶²	Chlorophyll-a, biomass, vegetation health
β_3 (difference vegetation index)	$\beta_3 \alpha (R_{rs}^{-1} (0.86) - R_{rs}^{-1} (0.66))$	Broge and Leblanc ⁴⁹	Vegetation greenness, chlorophyll-a, biomass
β_4	$\beta_4 \alpha (R_{rs}^{-1} (0.66))$	Huete et al. ⁶³	Chlorophyll-a, vegetation greenness
β_5	$\beta_5 \alpha (R_{rs}^{-1} (0.86))$	Huete et al. ⁶³	Vegetation greenness, biomass, moisture content, chlorophyll-a

Table 5: The results of the logistic regression as applied on April 2014, May 2015 and July 2016 Landsat 8 OLI images of the study area

Image date	Variable	Estimate	Standard error	Pr (> z)	κ	D^2 (%)	Overall accuracy % (Landsat 8/MODIS)
22 April 2014	y-intercept	22.7800	7.9400	0.0040**	0.74	76.38	80/95
	β_4	-0.0025	0.0001	0.0003***			
	β_5	0.0001	0.0004	0.8900			
27 May 2015	y-intercept	6.9939	3.1857	0.0281*	0.30	35.01	65/65
	β_3	0.0027	0.0011	0.0181*			
16 July 2016	y-intercept	1.090	1.419	0.442	0.43	64.37	83/-
	β_3	36.308	16.628	0.029*			

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Results and discussion

The results of the SLR are shown in Table 5. The SLR model used to estimate classes of algae based on chl-a concentration was more sensitive for the April 2014 analysis than for both May 2015 and 2016 data sets. For the April 2014 Landsat data, the SLR model yielded a higher D^2 than for the May 2015 data ($D^2 = 76.8\%$ and 35.01% , respectively). The April period in South Africa corresponds to the period when the algae and cyanobacteria in the Vaal Dam reach their maximum photosynthetic activities.³⁷ Classification of the 2016 image yielded the second highest D^2 of 64.37% . The equations used to derive chl-a distribution maps (classes) for both April 2014 and May 2015 are given by:

$$\text{chl-a} = \frac{\exp^{(22.7800 - (0.0025 \times \beta_4) + (0.0001 \times \beta_5))}}{(1 + \exp^{(22.7800 - (0.0025 \times \beta_4) + (0.0001 \times \beta_5))})} \quad \text{Equation 8}$$

$$\text{chl-a} = \frac{\exp^{(6.9989 + (0.0027 \times \beta_3))}}{(1 + \exp^{(6.9989 + (0.0027 \times \beta_3))})} \quad \text{Equation 9}$$

Equations 8 and 9 were applied to the 2014 and 2015 images, respectively. The model which comprises variables β_4 (red band at $0.66 \mu\text{m}$) and β_5 (NIR band at $0.86 \mu\text{m}$) as predictors was used for estimating chl-a concentration in the Vaal Dam. The red band (β_4) was significantly and negatively correlated with the density of chl-a in the dam (-0.002 ; $p < 0.001$) for the April 2014 Landsat 8 data. This relationship is as a result of the photochemical activity of photosystem II which is characterised by a reduction in the shorter wavelength red band because of increased chl-a concentration.⁴⁶ Additionally, although not significantly, the NIR band (β_5) was positively correlated with the density of chl-a (0.0001 ; $p < 0.89$).

The positive correlation between the chl-a and the NIR also supports the negative correlation between the red band and chl-a density. The increase in algae biomass results in an increase in the opposing trends between the NIR and red band correlations (Figure 3). The NIR region, which is outside of the PAR region, is mainly used by the algae for photomorphogenesis, which is light regulating changes in development, morphology, biochemistry, cell structure and function in response to light.^{47,48}

For the 2016 winter image, the estimation of chl-a was derived following Equation 10. The pattern of this model used for predicting chl-a concentration showed similarity to the one used in 2015.

$$\text{chl-a} = \frac{\exp^{(1.090 + (36.308 \times \beta_3))}}{(1 + \exp^{(1.090 + (36.308 \times \beta_3))})} \quad \text{Equation 10}$$

On the other hand, the difference vegetation index represented as variable β_3 was the only significant (0.0027 ; $p < 0.018$) remote sensing variable for predicting chl-a in the Vaal Dam for the images acquired in both May 2015 and July 2016. The positive correlation between chl-a classes and difference vegetation index suggests that the higher the concentration of the healthy green algae, the greater the absorption of the red band (PAR), as the difference vegetation index is calculated from the difference between the red and NIR bands.⁴⁹ It is generally known that remote sensing indices are better designed for mapping changes in chl-a than are individual spectral bands because of their ability to compensate for errors resulting from solar or viewing geometry.

Initial desktop validation data showed that the SLR model applied on Landsat OLI April 2014 data yielded an 80% overall accuracy while a 65% overall accuracy was obtained from Landsat OLI data for May 2015.

There was substantial agreement ($\kappa=0.74$) between the validation data set and the predicted chl-a concentrations for the image acquired on 22 April 2014, while a fair agreement ($\kappa=0.30$) was observed for the 22 May 2015 image. These levels of agreement were predominantly affected by the sample size of the validation data set. However, the field data set resulted in a classification accuracy of 83% with fair agreement between the observed and the measured concentrations ($\kappa=0.43$).

Thematic maps

Figure 5 shows the results of the spatial characteristics of predicted chl-a for the April 2014 and May 2015 study periods. The chl-a concentration ranged from 0 (indicating low class) to 0.99 (indicating high class) in the lentic freshwater body corresponding to a minimum of 2.5 $\mu\text{g/L}$ and a maximum of 1219 $\mu\text{g/L}$. On the other hand, there is correlation ($R^2=0.61$) between the log-transformed chl-a measurements and the estimated chl-a as shown in Figure 6. The resultant maps indicate that higher class concentrations were predominantly found at or near the shores of the dam. These areas are usually shallower with higher temperatures than the deep waters, resulting in lateral diurnal thermal variations.⁵⁰ In addition, the various concentrations of algal blooms are also attributed to the global changes in climate and anthropogenic activities.⁵¹ Whereas climate change affects algal formation at a global scale, the local anthropogenic activities – such as nutrient load into the Vaal Dam – are more important in controlling algal densities. This is particularly important considering the various forms of human activities occurring at different parts of the Vaal River Basin. The higher concentrations were usually found where the Dam has a meandering shape, especially in the northern part of the Dam, although some higher chl-a concentrations occurred within the Dam. The northern part is where the Vaal River flows into the Vaal Dam, contributing a greater proportion of nutrient load.⁵²

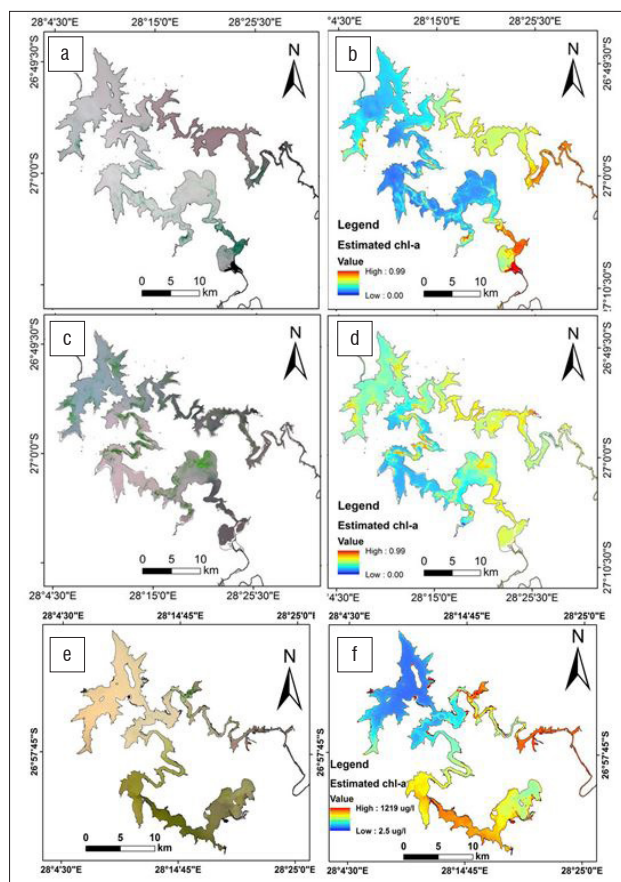


Figure 5: The stepwise logistic regression results depicting predicted distribution of chl-a within the Vaal Dam. The true colour maps are indicated by (a), (c) and (e) and their corresponding results in (b), (d) and (f).

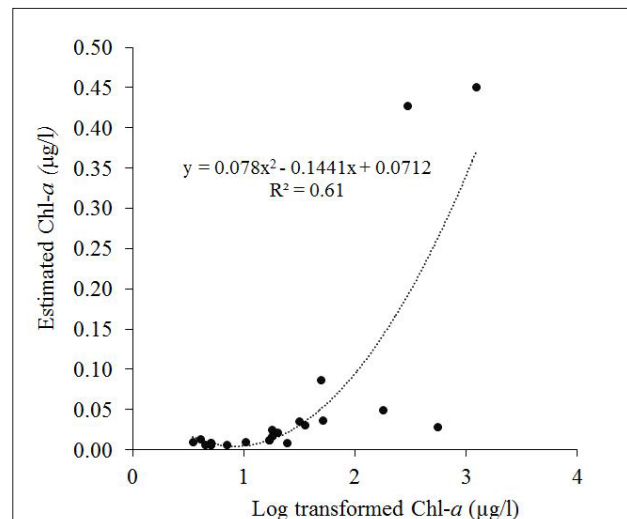


Figure 6: Correlation between the predicted chl-a concentrations and measured log-transformed chl-a concentrations within the Vaal Dam.

From the Landsat data acquired for April 2014, high concentrations of chl-a were estimated to be in the southeasterly part of the Dam – a flood plain composed mainly of productive agricultural lands. The sections located next to the Dam wall (in the northwestern side) had relatively low chl-a concentrations in 2014, while in 2015 there was a shift in terms of chl-a in that previously low class water sections became high class sections. From the 2016 image analysis, it becomes apparent that the 2015/2016 drought reported in South Africa resulted in increased algal blooms at the study area, especially where the Vaal River enters the Dam. There has been a pronounced shift in chl-a concentrations from very sparse concentrations to very dense concentrations since 2014. Factors that might have contributed to this spatial chlorophyll shift may be a response to the rapid warming of the water along Vaal tributaries during the September to November months, and rapid cooling during April to June.⁵³ These rapid temperature variations create a situation in which the growth rates of freshwater eukaryotic phytoplankton generally stabilise, while growth rates of many cyanobacteria increase, thereby providing a competitive advantage.^{54,55} The rapid growth of cyanobacteria in the Vaal Dam during April marks the final dominance of both *Anabaena* and *Microcystis* species, followed by a period of very low blooms during the winter months (May–August). The stability of the Vaal Dam (as a slow-circulating dam) and the availability of phosphates in the water are usually amongst the contributing factors to the sudden algal development in case II water bodies.⁵⁶ Both phosphorus and nitrogen are some of the crucial biochemical components of plant organic matter, and previous studies have shown that remote sensing data can be used to estimate such chemical components in plants.⁵⁷

Pollution in the Vaal Dam

Both cyanobacteria and algae grow naturally in water bodies. However, anthropogenic activities play a major role by influencing the rate of growth of algal blooms.^{6,58} Several anthropogenic activities were identified as primary contributors to the algal blooms in the Vaal Dam, resulting in an obvious increase in the eutrophication levels. These activities include dissolved urban effluents, heavy metal contaminants, mining and industrial effluent, most of which are nutrient rich.⁵⁹ These contaminants have their origin in tributaries upstream, such as those of the Vaal River and the Wilge River. The tributaries of the Vaal River may be most important because they drain from some highly industrialised areas of Gauteng and Mpumalanga. The concentrations of chl-a in the Vaal Dam have exceeded 70 $\mu\text{g/L}$ since 2005 and continue to increase annually as a result of a combination of many factors.⁶⁰ It has been reported that most of the significant water quality challenges emerge from biological materials (from faecal solid materials) and chemical materials (from gold mining and industrial pollutants).⁶¹ All of these factors render the Vaal Dam eutrophic, which means that the reservoir contains, among other water quality indicators, a chl-a greater than 5 $\mu\text{g/m}^3$.³⁷

Considerations

The aim of the study was to estimate the chl-a concentrations within the Vaal Dam, using the Landsat OLI data set. This study was successful in classifying chl-a into two classes – low concentrations and high concentrations. The SLR model used identified different significant variables within each data set, although the trend of occurrence was similar in all instances. The classification was done in autumn, which is the period that corresponds with the highest chl-a content within the Vaal Dam.³⁷ In the current study, high chl-a concentrations were associated with the presence of dominant Cyanophyceae group (*Anabaena* and *Mycrocystis*) species. The current analysis is primarily relevant for remediation purposes, in that it provides estimates of the status quo during the high cyanobacteria and algae blooming season. Mapping chl-a concentrations during these time periods is key for implementing remediation strategies against eutrophication as dictated by the South African *National Water Act (Act no. 36 of 1998)*, and does not guarantee similar concentrations in subsequent years unless the anthropogenic and biophysical/chemical conditions are maintained. Predicting the potential distribution of chl-a in the Vaal Dam during low cyanobacteria and algae periods could be essential for planning and management purposes. Knowing where the potentially high densities of chl-a are likely to be will enable the water resource managers to strategically allocate limited financial resources for water treatment using algacides in the Vaal Dam.

The mapping of chl-a in the current study may have been limited by the atmospheric correction method used for the study. It was not the purpose of the current study to compare the performance of atmospheric correction algorithms as such studies have been done elsewhere.¹¹ However, it is possible that the accuracy of estimates may have been affected by the atmospheric correction method used. Additionally, the number of training and validation points may have impacted on the accuracy of the SLR model for both Landsat OLI and MODIS data sets. The difference in ground sampling distance between Landsat 8 OLI (15 m pan-sharpened) and MODIS (250 m) could have affected the classification accuracy of the SLR model, as could the lower number of validation points ($n=20$). However, apart from these considerations, Landsat 8 has shown potential for mapping chl-a concentrations in turbid waters in South Africa.

Conclusions

Both the remote sensing indices (difference vegetation index) and the individual spectral bands (NIR-red) were used successfully to estimate chl-a in the turbid Vaal Dam water. The sensitivity of NIR-red models is dependent upon the optical characteristics of a reservoir and thus has different variable significance in similar locations but at different times. The field data set used for the 2016 Landsat 8 image showed that there exists a strong correlation between predicted and measured chl-a concentrations peculiar to the Vaal Dam. Landsat 8 data remain useful for estimation of chl-a concentrations in trophic waters and the integration of information from both MODIS and Landsat coupled with field data could assist in constant monitoring of larger water bodies in South Africa. The Landsat heritage is crucial for studying long-term variations of slow-circulating water bodies in South Africa amidst water-related challenges faced nationally. Additionally, testing the effect of atmospheric correction algorithms for South African conditions could be useful, although the QUAC atmospheric correction algorithm used produced reliable results with a high classification accuracy.

Acknowledgements

We thank the Earth Observation Division of the South African National Space Agency (SANSA-EO) for the support required to write the manuscript. We also thank Ms Nale Mudau of SANSA-EO, and Mr Reveck Hariram and Dr Annelie Swanepoel of Rand Water for assisting in validation data logistics, collection and laboratory analysis. We acknowledge the USGS for making Landsat and MODIS data readily available to be used for this study. We thank the anonymous reviewers for improving the quality and readability of this paper.

Authors' contributions

O.E.M. and T.O. conceptualised and designed the study. O.E.M. and L.T. collected the data used in the study. O.E.M., T.O. and L.T.T. analysed the data. P.M. was the project leader and was responsible for the budget. O.M. and P.M. wrote, reviewed, edited and approved the manuscript. O.E.M., T.O., L.T.T. and P.M. edited and approved the revised manuscript.

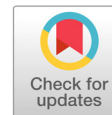
References

- Zhang Z, Mei Z. Effects of human activities on the ecological changes of lakes in China. *GeoJournal*. 1996;40(1–2):17–24. <https://doi.org/10.1007/BF00222526>
- Khan M. Pollution of water resources due to industrialization in arid zone of Rajasthan, India. *J Environ Sci*. 2001;13(2):218–223.
- Mulholland PJ. Assessment and control of nonpoint source pollution of aquatic ecosystems. A practical approach. *Limnol Oceanogr*. 2001;46(1):212–212.
- Obi C, Potgieter N, Bessong P, Matsaung G. Assessment of the microbial quality of river water sources in rural Venda communities in South Africa. *Water SA*. 2002;28(3):287–292. <https://doi.org/10.4314/wsa.v28i3.4896>
- Rivett U, Champanis M, Wilson-Jones T. Monitoring drinking water quality in South Africa: Designing information systems for local needs. *Water SA*. 2013;39(3):409–414. <https://doi.org/10.4314/wsa.v39i3.10>
- Van Ginkel C. Eutrophication: Present reality and future challenges for South Africa. *Water SA*. 2011;37(5):693–701. <https://doi.org/10.4314/wsa.v37i5.6>
- Majozi NP, Salama MS, Bernard S, Harper DM, Habte MG. Remote sensing of euphotic depth in shallow tropical inland waters of Lake Naivasha using MERIS data. *Rem Sens Environ*. 2014;148:178–189. <https://doi.org/10.1016/j.rse.2014.03.025>
- Papoutsas C, Retalis A, Toulous L, Hadjimitsis DG, editors. Monitoring water quality parameters for Case II waters in Cyprus using satellite data. *Proc SPIE*. 2014:9229. <https://doi.org/10.1117/12.2069788>
- Dube T, Mutanga O, Seutloali K, Adelabu S, Shoko C. Water quality monitoring in sub-Saharan African lakes: A review of remote sensing applications. *Afr J Aqua Sci*. 2015;40(1):1–7. <https://doi.org/10.2989/16085914.2015.1014994>
- Brando VE, Dekker AG. Satellite hyperspectral remote sensing for estimating estuarine and coastal water quality. *IEEE Trans Geosci Rem Sens*. 2003;41(6):1378–1387. <https://doi.org/10.1109/TGRS.2003.812907>
- Moses WJ, Gitelson AA, Berdnikov S, Povazhnyy V. Estimation of chlorophyll-a concentration in case II waters using MODIS and MERIS data—successes and challenges. *Environ Res Letters*. 2009;4(4), Art. #045005, 8 pages. <https://doi.org/10.1088/1748-9326/4/4/045005>
- Matthews MW, Bernard S, Winter K. Remote sensing of cyanobacteria-dominant algal blooms and water quality parameters in Zeekoevlei, a small hypertrophic lake, using MERIS. *Rem Sens Environ*. 2010;114(9):2070–2087. <https://doi.org/10.1016/j.rse.2010.04.013>
- Becker BL, Lusch DP, Qi J. A classification-based assessment of the optimal spectral and spatial resolutions for Great Lakes coastal wetland imagery. *Rem Sens Environ*. 2007;108(1):111–120. <https://doi.org/10.1016/j.rse.2006.11.005>
- Hestir EL, Khanna S, Andrew ME, Santos MJ, Viers JH, Greenberg JA, et al. Identification of invasive vegetation using hyperspectral remote sensing in the California Delta ecosystem. *Rem Sens Environ*. 2008;112(11):4034–4047. <https://doi.org/10.1016/j.rse.2008.01.022>
- Dlamini S, Nhapi I, Gumindoga W, Nhwatiwa T, Dube T. Assessing the feasibility of integrating remote sensing and in-situ measurements in monitoring water quality status of Lake Chivero, Zimbabwe. *Phys Chem Earth*. 2016;93:2–11. <https://doi.org/10.1016/j.pce.2016.04.004>
- Lee Z, Carder KL, Mobley CD, Steward RG, Patch JS. Hyperspectral remote sensing for shallow waters 2: Deriving bottom depths and water properties by optimization. *Appl Opt*. 1999;38(18):3831–3843. <https://doi.org/10.1364/AO.38.003831>
- Randolph K, Wilson J, Tedesco L, Li L, Pascual DL, Soyeyx E. Hyperspectral remote sensing of cyanobacteria in turbid productive water using optically active pigments, chlorophyll a and phycocyanin. *Rem Sens Environ*. 2008;112(11):4009–4019. <https://doi.org/10.1016/j.rse.2008.06.002>

18. Gitelson AA, Gao B-C, Li R-R, Berdnikov S, Saprygin V. Estimation of chlorophyll-a concentration in productive turbid waters using a Hyperspectral Imager for the Coastal Ocean—the Azov Sea case study. *Environ Res Letters*. 2011;6(2), Art. #024023, 6 pages. <https://doi.org/10.1088/1748-9326/6/2/024023>
19. Ha NTT, Koike K, Nhuon MT. Improved accuracy of chlorophyll-a concentration estimates from MODIS imagery using a two-band ratio algorithm and geostatistics: As applied to the monitoring of eutrophication processes over Tien Yen Bay (Northern Vietnam). *Rem Sens*. 2013;6(1):421–442. <https://doi.org/10.3390/rs6010421>
20. Dall’Olmo G, Gitelson AA, Rundquist DC, Leavitt B, Barrow T, Holz JC. Assessing the potential of SeaWiFS and MODIS for estimating chlorophyll concentration in turbid productive waters using red and near-infrared bands. *Rem Sens Environ*. 2005;96(2):176–187. <https://doi.org/10.1016/j.rse.2005.02.007>
21. Chawira M, Dube T, Gumindoga W. Remote sensing based water quality monitoring in Chivero and Manyame Lakes of Zimbabwe. *Phys Chem Earth*. 2013;66:38–44. <https://doi.org/10.1016/j.pce.2013.09.003>
22. Kallio K, Attila J, Härmä P, Koponen S, Pulliainen J, Hyytiäinen U-M, et al. Landsat ETM+ images in the estimation of seasonal lake water quality in boreal river basins. *Environ Manage*. 2008;42(3):511–522. <https://doi.org/10.1007/s00267-008-9146-y>
23. Gerace AD, Schott JR, Nevins R. Increased potential to monitor water quality in the near-shore environment with Landsat’s next-generation satellite. *J Appl Rem Sens*. 2013;7(1), Art. #073558, 19 pages. <https://doi.org/10.1117/1.JRS.7.073558>
24. Yin G, Mariethoz G, McCabe MF. Gap-filling of Landsat 7 imagery using the direct sampling method. *Rem Sens*. 2016;9(1), Art. #12, 20 pages. <https://doi.org/10.3390/rs9010012>
25. Torbick N, Corbiere M. A multiscale mapping assessment of Lake Champlain cyanobacterial harmful algal blooms. *Int J Environ Res Pub Health*. 2015;12(9):11560–11578. <https://doi.org/10.3390/ijerph120911560>
26. Gitelson AA, Dall’Olmo G, Moses W, Rundquist DC, Barrow T, Fisher TR, et al. A simple semi-analytical model for remote estimation of chlorophyll-a in turbid waters: Validation. *Rem Sens Environ*. 2008;112(9):3582–3593. <https://doi.org/10.1016/j.rse.2008.04.015>
27. Dall’Olmo G, Gitelson AA. Effect of bio-optical parameter variability on the remote estimation of chlorophyll-a concentration in turbid productive waters: Experimental results. *Appl Opt*. 2005;44(3):412–422. <https://doi.org/10.1364/AO.44.000412>
28. Hoge FE, Lyon PE. Satellite retrieval of inherent optical properties by linear matrix inversion of oceanic radiance models: An analysis of model and radiance measurement errors. *J Geophys Res Oceans*. 1996;101(C7):16631–16648. <https://doi.org/10.1029/96JC01414>
29. Schiller H, Doerffer R. Neural network for emulation of an inverse model operational derivation of Case II water properties from MERIS data. *Int J Rem Sens*. 1999;20(9):1735–1746. <https://doi.org/10.1080/014311699212443>
30. Chang N-B, Daranpob A, Yang YJ, Jin K-R. Comparative data mining analysis for information retrieval of MODIS images: Monitoring lake turbidity changes at Lake Okeechobee, Florida. *J Applied Rem Sens*. 2009;3(1):033549. <https://doi.org/10.1117/1.3244644>
31. Zimba PV, Gitelson A. Remote estimation of chlorophyll concentration in hyper-eutrophic aquatic systems: Model tuning and accuracy optimization. *Aquaculture*. 2006;256(1–4):272–286. <https://doi.org/10.1016/j.aquaculture.2006.02.038>
32. Gouws K, Coetzer P. Determination and partitioning of heavy metals in sediments of the Vaal Dam System by sequential extraction. *Water SA*. 1997;23:217–226.
33. Department of Water Affairs and Forestry. Vaal Dam [webpage on the Internet]. No date [cited 2018 Aug 14]. Available from: <http://www.dwaf.gov.za/orange/Vaal/vaaldam.htm>
34. RandWater. Where does our water come from? The Vaal River System [webpage on the Internet]. No date [cited 2016 Jul 21]. <http://www.waterwise.co.za/site/water/purification/>
35. Cloot A, Le Roux G. Modelling algal blooms in the middle Vaal River: A site specific approach. *Water Res*. 1997;31(2):271–279. [https://doi.org/10.1016/S0043-1354\(96\)00248-5](https://doi.org/10.1016/S0043-1354(96)00248-5)
36. Kellerman T, Coetzer J, Naudé T, Botha C. Plant poisonings and mycotoxicoses of livestock in southern Africa. New York: Oxford University Press Southern Africa;2005.
37. Matthews MW, Bernard S. Eutrophication and cyanobacteria in South Africa’s standing water bodies: A view from space. *S Afr J Sci*. 2015;111(5–6), Art. #2014-0193, 8 pages. <https://doi.org/10.17159/sajs.2015/20140193>
38. Malahlela OE. Inland waterbody mapping: Towards improving discrimination and extraction of inland surface water features. *Int J Rem Sens*. 2016;37(19):4574–4589. <https://doi.org/10.1080/01431161.2016.1217441>
39. Exelis Visual Information Solutions. Environment for visualizing images. Boulder, CO: Exelis Visual Information Solutions; 2017.
40. R Core Development Team. The R project for statistical computing. Version 3.1. Vienna: R Foundation; 2017.
41. Kganyago M, Odindi J, Adjorlolo C, Mhangara P. Evaluating the capability of Landsat 8 OLI and SPOT 6 for discriminating invasive alien species in the African savanna landscape. *Int J Appl Earth Obs Geoinf*. 2018;67:10–19. <https://doi.org/10.1016/j.jag.2017.12.008>
42. Deng Y, Wu C, Li M, Chen R. RNDI: A ratio normalized difference soil index for remote sensing of urban/suburban environments. *Int J Appl Earth Obs Geoinf*. 2015;39:40–48. <https://doi.org/10.1016/j.jag.2015.02.010>
43. Geng L, Ma M, Yu W, Wang X, Jia S. Validation of the MODIS NDVI products in different land-use types using in situ measurements in the Heihe River Basin. *IEEE Geosci Rem Sens Lett*. 2014;11(9):1649–1653. <https://doi.org/10.1109/LGRS.2014.2314134>
44. Gitelson AA, Peng Y, Huemmrich KF. Relationship between fraction of radiation absorbed by photosynthesizing maize and soybean canopies and NDVI from remotely sensed data taken at close range and from MODIS 250 m resolution data. *Rem Sens Environ*. 2014;147:108–120. <https://doi.org/10.1016/j.rse.2014.02.014>
45. Hashemi SA, Chenani SK. Investigation of NDVI index in relation to chlorophyll content change and phenological event. *Rec Adv Environ Eng Syst Nav Sci*. 2011:22–28.
46. Pettai H, Oja V, Freiberg A, Laisk A. Photosynthetic activity of far-red light in green plants. *Biochim Biophys Acta Bioenerg*. 2005;1708(3):311–321. <https://doi.org/10.1016/j.bbabi.2005.05.005>
47. Nishio J. Why are higher plants green? Evolution of the higher plant photosynthetic pigment complement. *Plant Cell Environ*. 2000;23(6):539–548. <https://doi.org/10.1046/j.1365-3040.2000.00563.x>
48. Han Y-J, Song P-S, Kim J-I. Phytochrome-mediated photomorphogenesis in plants. *J Plant Biol*. 2007;50(3):230–240. <https://doi.org/10.1007/BF03030650>
49. Broge NH, Leblanc E. Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. *Rem Sens Environ*. 2001;76(2):156–172. [https://doi.org/10.1016/S0034-4257\(00\)00197-8](https://doi.org/10.1016/S0034-4257(00)00197-8)
50. Tonolla D, Acuna V, Uehlinger U, Frank T, Tockner K. Thermal heterogeneity in river floodplains. *Ecosystems*. 2010;13(5):727–740. <https://doi.org/10.1007/s10021-010-9350-5>
51. Clark JM, Schaeffer BA, Darling JA, Urquhart EA, Johnston JM, Ignatius AR, et al. Satellite monitoring of cyanobacterial harmful algal bloom frequency in recreational waters and drinking water sources. *Ecol Indic*. 2017;80:84–95. <https://doi.org/10.1016/j.ecolind.2017.04.046>
52. Gyedu-Ababio TK, Van Wyk F. Effects of human activities on the Waterval River, Vaal River catchment, South Africa. *Afr J Aqua Sci*. 2010;29(1):75–81.
53. Roos JC, Pieterse A. Light, temperature and flow regimes of the Vaal River at Balkfontein, South Africa. *Hydrobiology*. 1994;277(1):1–15. <https://doi.org/10.1007/BF00023982>
54. Peperzak L. Climate change and harmful algal blooms in the North Sea. *Acta Oecol*. 2003;24:S139–S144. [https://doi.org/10.1016/S1146-609X\(03\)00009-2](https://doi.org/10.1016/S1146-609X(03)00009-2)
55. Paerl HW, Huisman J. Climate change: A catalyst for global expansion of harmful cyanobacterial blooms. *Environ Microbiol Rep*. 2009;1(1):27–37. <https://doi.org/10.1111/j.1758-2229.2008.00004.x>

56. Schindler DW, Carpenter SR, Chapra SC, Hecky RE, Orihel DM. Reducing phosphorus to curb lake eutrophication is a success. *Environ Sci Technol*. 2016;50(17):8923–8929. <http://dx.doi.org/10.1021/acs.est.6b02204>
57. Zhai Y, Cui L, Zhou X, Gao Y, Fei T, Gao W. Estimation of nitrogen, phosphorus, and potassium contents in the leaves of different plants using laboratory-based visible and near-infrared reflectance spectroscopy: Comparison of partial least-square regression and support vector machine regression methods. *Int J Rem Sens*. 2013;34(7):2502–2518. <https://doi.org/10.1080/01431161.2012.746484>
58. Kings S. Sewage in Gauteng's drinking water. *Mail & Guardian*. 2015 July 24; Environment. Available from: <http://mg.co.za/article/2015-07-23-sewage-in-gautengs-drinking-water>
59. Kilner K, Hugo B. Save the Vaal... again!: *Environmental Law Without Prejudice*. 2015;15(8):22–24.
60. Chinyama A, Ochieng GM, Snyman J, Nhapi I. Occurrence of cyanobacteria genera in the Vaal Dam: Implications for potable water production. *Water SA*. 2016;42(3):415–420. <https://doi.org/10.4314/wsa.v42i3.06>
61. Du Plessis A. Primary water quality challenges for South Africa and the Upper Vaal WMA. *Freshwater challenges of South Africa and its Upper Vaal River*. Zug: Springer; 2017 p. 99–118. https://doi.org/10.1007/978-3-319-49502-6_6
62. Jordan CF. Derivation of leaf area index from quality of light on the forest floor. *Ecology*. 1969;50:663–666.
63. Huete A, Didan K, Miura T, Rodriguez EP, Gao X, Ferreira LG. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Rem Sens Environ*. 2002;83:195–213.





Application of science and technology by the South African food and beverage industry

AUTHORS:

Lisa-Claire Ronquest-Ross¹

Nick Vink²

Gunnar O. Sigge¹

AFFILIATIONS:

¹Department of Food Science, Stellenbosch University, Stellenbosch, South Africa

²Department of Agricultural Economics, Stellenbosch University, Stellenbosch, South Africa

CORRESPONDENCE TO:

Lisa-Claire Ronquest-Ross

EMAIL:

lisaronquest@gmail.com

DATES:

Received: 16 Aug. 2017

Revised: 09 Jan. 2018

Accepted: 02 June 2018

Published: 11 Sep. 2018

KEYWORDS:

South Africa; food and beverage manufacturing industry; advances in science and technology; packaged food and beverages

HOW TO CITE:

Ronquest-Ross L-C, Vink N, Sigge GO. Application of science and technology by the South African food and beverage industry. *S Afr J Sci.* 2018;114(9/10), Art. #4757, 11 pages. <https://doi.org/10.17159/sajs.2018/4757>

ARTICLE INCLUDES:

× Supplementary material

× Data set

FUNDING:

None

Significant shifts in the type of foods consumed by South Africans have taken place since 1994 and packaged food and beverage innovation has accelerated since then. Globally, advances in science and technology have benefitted food processing and food manufacturing technologies and systems. Significant capital investments have been made by the South African food and beverage manufacturing industry (SAFBMI). It is, however, not clear which technology areas have received investments and for what purposes. The objective of this study was thus to understand how the SAFBMI has invested in and applied science and technology since 1994. Data were sourced from food and beverage trade magazines, dating from 1986 to 2012. Trends over the past 30 years were analysed to determine the application of science and technology. The findings suggest that the dairy, soft drinks and bakery sectors have been most active. The main advances were to upgrade manufacturing facilities and build new plants to increase capacity, deliver new products and improve efficiencies and product quality and safety. Investments to improve thermal processing and packaging were also noted. We found evidence of the application of commercially available new preservation technologies and a low level of experimentation with non-commercial novel technologies by the SAFBMI. South Africa appears to be keeping pace with advances in food manufacturing in automation, process control and quality and food safety practices, material handling, and centralised distribution centres with warehouse management systems. Continued investment in food science and technology research will ensure that the growing consumer demand for packaged foods and beverages is met.

Significance:

- Although South Africa is keeping pace with science and technology advances in its food and beverage manufacturing operations, the need to continue to conduct and apply science and technology research, especially in novel processing, is highlighted.

Introduction

The South African context

South Africa changed after 27 April 1994 – changes that have impacted food consumption patterns as a result of shifts in food availability, accessibility and choices. Post-apartheid economic and transformation plans resulted in an increase in per capita income as well as the rise of a black middle class.¹ However, population growth, together with urbanisation resulting from the removal of pass laws and disbanding of homelands, as well as an increase in women entering the labour market, have resulted in high unemployment levels. This unemployment rate is despite an increase in actual numbers of people employed since 1994, with youth unemployment remaining a real concern.¹ There has also been significant growth of supermarkets, accounting for about 50–60% of retail sales.² Access to water, sanitation and electricity have all advanced.¹ Electricity allows access to refrigerators, ovens and microwaves, offering alternative food choices. The South African National Health and Nutrition Examination Survey-1 revealed that almost half (48%) of South Africans reported that they had eaten out before, with 28.3% of South Africans eating out weekly.³

The South African food and beverage manufacturing industry

The phenomenon of globalisation is having a major impact on food systems around the world. Food systems are changing, resulting in greater availability and diversity of food.⁴ Over the last 50 years, practically all areas of food production and processing have been transformed.⁵ This increase has resulted in a consolidation of agricultural and food and beverage companies into large international corporations, which have developed global brands and marketing strategies with adaptation to local tastes.^{5,6} These large corporations utilise global sourcing of supplies, centralisation of strategic assets and resources, as well as establishment of operations in several countries to succeed in a global market.^{6,7}

In 2012, there were over 1800 food manufacturing companies in South Africa.⁸ The South African food and beverage manufacturing industry (SAFBMI) is dominated by a few large, diversified, national and multinational food manufacturers, which have established market shares and control both production capacity and sales in most food categories.⁹ The ten largest packaged-food companies in South Africa account for 52% of total packaged food sales.¹⁰ This proportion is greater than the global average, for which the top ten packaged food companies account for only 15.2% of sales.¹⁰ Furthermore, this figure is different from that of the soft drinks sector, where the top ten global and South African soft drink companies account for 52.3% and 79% of sales, respectively.¹⁰ Five of the top ten food manufacturing companies in South Africa in 2012 were South African, three of which had an international presence.^{8,11} The domination of a few large food and beverage manufacturers is a consequence of restricted licensing procedures and technical barriers to entry that limited the number of manufacturers under the apartheid government.¹² However, more recently, new and smaller processors have started to play an important role and

continue to change the competitive environment of food processing in South Africa despite their limited market share.⁹

Large multinational companies such as Nestlé (Switzerland), Unilever (United Kingdom/Netherlands), Lactalis/Parmalat Group (France/Italy) and Mondelez (United States of America) operate their own manufacturing plants in South Africa, which produce and market their established brands.⁹ South Africa's key local players include companies such as Tiger Brands Ltd, Clover Ltd, Pioneer Foods Group Ltd, AVI Ltd, Premier Foods and Rainbow Chicken Limited. All of these food and beverage manufacturers depend on formal retail chains to sell their products.⁹ Many South African companies have formed associations with international companies, enabling access to the latest technology and expertise. Examples of these are Simba and Frito-Lay (United States of America) and Robertsons and Best Foods (United States of America).⁹

According to the Euromonitor 2012 report on the South African packaged food sector, this sector has grown from 2007 to 2012, in value by 57%, from ZAR91 billion to ZAR143 billion, and in volume by 15%, from 4 515 300 tonnes to 5 202 500 tonnes.¹³ Capital expenditure on new property, plants and equipment increased by 148%, from ZAR4010 million in 2001 to ZAR9943 million in 2008.^{14,15}

Food and beverage innovation in South Africa

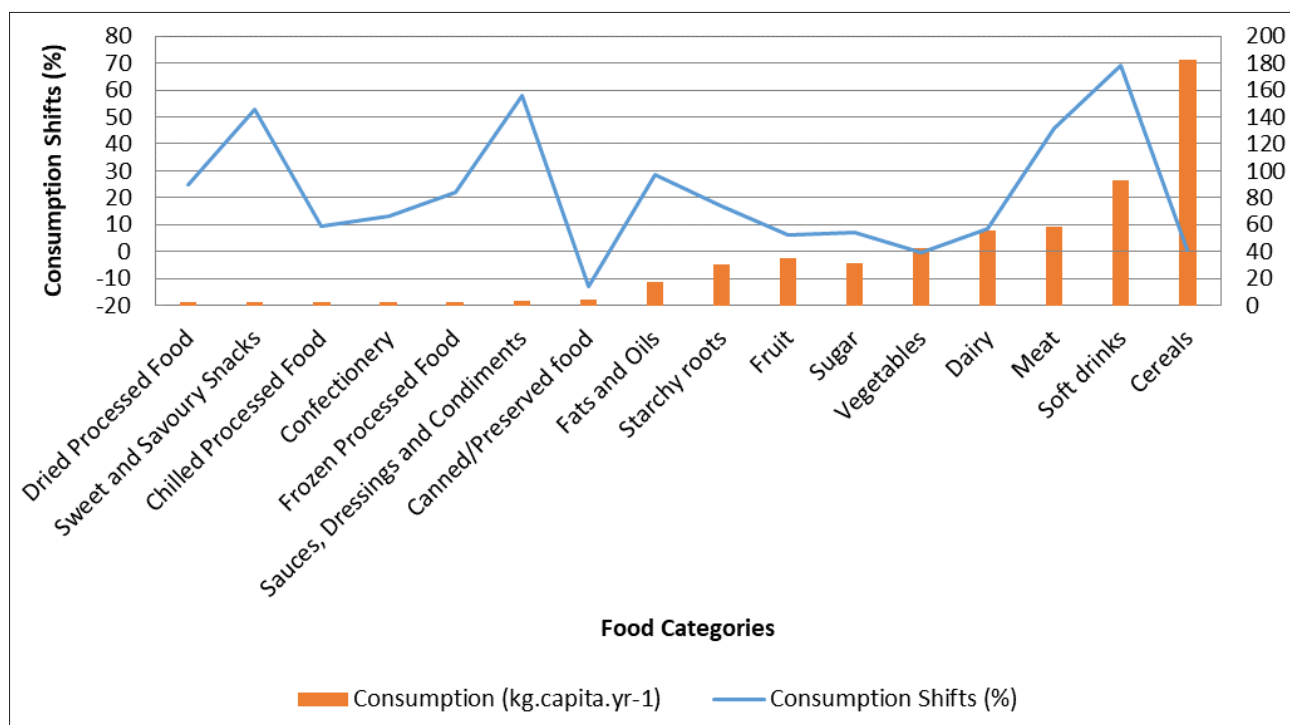
In recent years, there has been an increase in sales of almost all categories of packaged foods and beverages in South Africa.^{8,13} South African data derived from the Innova database indicates there has been a significant increase in product innovations and introductions in South Africa in the packaged food and beverage category since 2008.¹⁶ The categories with the highest number of product innovations and introductions were confectionery, sauces and seasonings, snacks, ready meals, soft drinks and hot drinks.¹⁶ There is a direct correlation between the development and implementation of new technologies and new, more desirable products on the shelf.^{17,18} The leading positioning behind these launches were convenience, health and choice.¹⁶ The

increase in product innovation has translated into increased consumption of packaged food across all categories, with a decline of 13% only in canned/preserved food.¹⁹

Food consumption changes since 1994

Food consumption patterns have changed and will continue to change dramatically over the coming decades in South Africa.¹⁹⁻²¹ Various food-related studies conducted over the past decades indicate that food consumption shifts in South Africa are towards a more Western-oriented diet with nutritional consequences contributing to increased obesity and other non-communicable diseases.²²⁻²⁴ A recent study by Ronquest-Ross et al.¹⁹ indicated that food consumption shifts are towards an overall increase in daily kilojoules consumed, a diet of sugar-sweetened beverages, increased proportion of processed and packaged food including edible vegetable oils, increased intake of animal-source foods, and added caloric sweeteners and away from vegetables (Figure 1). The largest shifts in food consumption were observed for soft drinks, sauces, dressings and condiments, sweet and savoury snacks, meat and fats and oils (Figure 1). These shifts in food consumption are concerning with respect to nutritional composition and public health.¹⁹ Convenience, health and wellness, and indulgence were the main consumer drivers for the increase in consumption of packaged foods since 1994.¹⁹

Locally and globally, fast food, bottled soft drink and multinational food companies are often implicated in the increase in non-communicable diseases.^{22,25} The Department of Health has therefore targeted the SAFBMI with regulations in an attempt to improve public health. These regulations include mandatory fortification of staple foods, maximum limit on trans-fat content in foodstuffs and reduction of salt in certain foodstuffs. Further regulations related to food composition and/or labelling as well as consumer demand for healthier and affordable products will require advances in and/or the application of scientific and technological developments by the SAFBMI.



Source: Food and Agriculture Organization of the United Nations Statistics Division Food Balance Sheets and Euromonitor Packaged Food and Beverage Consumption. Taken from Ronquest-Ross et al.¹⁹

Figure 1: Per capita consumption and shifts in consumption of specific food categories in South Africa from 1994/1999 to 2009/2012.

Advances in food processing technologies and manufacturing

Worldwide advances in food processing and packaging technologies

Advancement in processing technologies through science and technology is one of the best tools to meet the challenges accompanying regulatory pressure to reduce the impact of food processing on the environment, meet the need for a stable and secure food supply and meet changing consumer demand for safer, better quality, 'fresher-tasting' or minimally processed foods.^{18,26} To meet these challenges, the food industry is devoting considerable resources and expertise to improving the way in which foods are produced, distributed, stored and retailed.²⁶

Food preservation techniques can be classified as traditional if they are based on (1) the effects of temperature, (2) reduction of water activity or pH or (3) addition of preservatives. Food preservation techniques can be classified as emerging or novel preservation technologies when based on other processes.²⁷ The main advantages of traditional techniques are lower operational costs, safety and the fact that they are well established and widely used.²⁷ Emerging or novel preservation technologies may be an important complement to existing traditional processes for certain foods, but their use is associated with certain risks which need to be evaluated by industries and regulatory authorities before commercialisation.²⁷ Before successful development, implementation and adoption of a novel technology can take place, consumer acceptance, product quality, preservation efficacy and final product safety need to be assessed.²⁸

There is extensive knowledge on thermal processing as a food preservation method and advances in technology enabled optimisation for maximum efficacy against microbial contamination with minimum deterioration of food quality.²⁶ However, there is still a need for technologies that minimise the destructive influence of heat on food.²⁹ Because thermal energy has to be transmitted across the product itself to ensure effective microbial deactivation, even concepts like high-temperature-short-time processing are limited by the physical properties of the product.²⁹ By 2012, there had been limited industrial applications of microwave heating, even though it is widely accepted by consumers for preparing foods and extensive research on this method has already been conducted.^{30,31} The process of ohmic heating involves passing alternating electrical current through a food product that serves as electrical resistance, causing it to heat up.^{31,32} There are a number of factors that affect the heating rate of foods with ohmic heating and therefore commercial applications are only for high-acid foods and these applications have been slow to realise.³²

Novel non-thermal processing is being explored intensively to resolve many of the shortcomings related to thermal treatment as well as to meet consumer demands for minimal processing. Promising non-thermal pasteurisation technologies exist, such as high pressure, high-intensity pulsed electric field, pulsed light, oscillating magnetic field pulses and electron ionising radiation, which may be combined to yield the total bacterial reduction desired through the 'hurdle concept' of applying gentle steps to gradually reduce microbial counts.^{30,33} Apart from high-pressure processing and pulsed electric field, by 2012, there had been limited commercial applications of other emerging non-thermal processing technologies because of technological barriers to standardise processes and validate product quality and safety.^{30,34-36}

Packaging plays an important role in the food manufacturing process by making packaged foods more convenient, extending shelf life, allowing distribution across wide geographies while maintaining quality and food safety as well as providing containment, marketing and information, traceability and tamper-evidence.^{17,37,38} In order to meet the huge demand for processed food with a longer shelf life, various new methods of packaging are being used in the food-processing industry.³⁷ Packaging can manage oxygen in the package through gas flushes, controlling internal atmosphere or modified atmosphere packaging and vacuum, and enabling clean-label products without preservatives.^{18,39,40} 'Smart packaging' is a term used to describe packaging that is either active or intelligent and provides information to the user through absorption or release of substances to or from the food or environment.^{40,41}

Nanotechnology has been hailed as the industrial revolution of the 21st century and has been applied to food packaging to improve barrier properties, and provide active antimicrobial and antifungal surfaces, mechanical and heat-resistance properties, sensing and signalling microbiological and biochemical changes, traceability purposes and monitoring and repairing of tears in packaging.^{38,40,42,43}

The beverage industry has always been at the forefront of adopting innovative new packaging technologies.⁴⁴ Examples of these are gas-release packaging through a 'widget', flavour-release packaging like the Unistraw™ and nutrient-, probiotic- or enzyme-release packaging.⁴⁴ Temperature control, such as in self-heating of plastic coffee cans or ready meals, thermochromic labelling to indicate the temperature of a beverage and improved tamper-proof packaging are other examples.^{38,44,45}

Packaging technology advances the need to balance food protection with numerous issues such as energy and material costs, growing social and environmental concerns, strict regulations on pollutants and disposal of municipal solid waste.^{17,46} Reducing, reusing and recycling are the main initiatives in reducing the environmental impact of packaging.⁴⁷ In terms of alternative packaging materials, biodegradable, renewable polymers are being extensively researched, with cellophane being the most common.⁴⁶

Advances in food manufacturing

Technology can be used in all parts of the production process of food-processing plants⁴⁸ and the competitive food manufacturer aims to adopt technology that will yield maximum efficiency and increased production throughputs at lower costs in their processing plants with better quality and enhanced food safety.^{28,40,49} Automation in the food industry has delivered improved productivity, product quality and profitability benefits.^{40,49} Furthermore, advances in materials handling and process control have led to substantial improvements in product quality, efficiency and cost reduction.⁵⁰ Important techniques in achieving efficient materials handling include continuous handling techniques as opposed to manual handling, bulk handling, automation, the use of gravity (where possible), combining operations to eliminate intermediate steps, conveyor systems and using a systems approach from raw material to finished product to avoid bottlenecks or shortages.^{40,50}

Advances in process control include the shift from controls reliant on operator skill to technology-based controls because of advances in microelectronics.⁵⁰ A significant development in process control was the introduction of programmable logic controllers in the 1980s, which allow instructions to be programmed and executed automatically based on data received by sensors.^{50,51} The advances in microelectronics together with computer software technology developments have allowed fast data processing, giving manufacturers more sophisticated, efficient, affordable, interlinked and user-friendly process control systems.^{50,51} These systems are applied throughout the manufacturing process, from material resource planning and production planning and management to control of process conditions including product quality, energy consumption and product flow, collation and evaluation of process and product data, control of cleaning-in-place and packaging, warehousing, storage and distribution control.⁵⁰ A further significant development software is called Supervisory Control and Data Acquisition, which collects data from the programmable logic controller and displays it as animated graphics so that trends and historical data reports can be generated and adjustments can be made in real time.^{49,50} Robotics are another advancement, which came about as a result of advances in neural networks, vision systems, pressure-sensitive grippers and laser guidance systems. This technology is becoming more of a permanent feature in food manufacturing facilities and has been applied for picking and placing items into packaging as well as palletising of loads.^{40,50,52}

In terms of warehousing and distribution, developments in software – such as radio-frequency identification and automation of the complete purchase process – have enabled significant improvements.^{40,50} Modelling and simulation analysis in the 1980s and 1990s by retailers to reduce logistics costs and distribution times led to the development of fewer, larger regional distribution centres that could handle a range

of products.^{50,53} The adoption of electronic data interchange allowed larger retailers to automatically replace products based on consumer purchases, effectively resulting in more frequent deliveries of small amounts to stores from regional distribution centres.⁵⁰

Advances in food safety and quality management by food manufacturers have been mainly driven by increased public and private standards. The control of food hazards (biological, chemical and physical) by food manufacturers has been through hazard analysis critical control points, which is a science-based approach to identify and establish control measures for specific hazards.^{32,40} Contaminants in products must be inspected and removed and technological advances such as electromagnetic spectrum X-ray or vision systems have enabled a non-destructive method with which to achieve this.^{40,49} Hyperspectral imaging is another inspection method for the analysis of the chemical composition of food products, allowing for examination of disease conditions, ripeness, hardness/tenderness, grading or contamination.^{49,54}

Hygiene control systems should be applied throughout the food chain and, more specifically, to proper product and process design from a food manufacturer perspective.³² Equipment design for ease and effectiveness of cleaning is another aspect on which much attention has been focused, such as cleaning-in-place and material selection for equipment.^{40,50} Being able to trace a food product from production to distribution is critical because of the consumer focus on quality with more complex supply chains. This traceability can be achieved through barcodes or radio-frequency identification.^{40,49}

A recent study of the United Kingdom food and beverage industry revealed that 55% of food and beverage manufacturers identified the primary technological innovation need as being related to improving product quality or delivering new products despite growing pressure on improving sustainability and resource efficiency.⁵⁵ Emerging technology trends included efficiency, productivity, sustainability and salt and fat reduction. Furthermore, 60% of these food and beverage manufacturers felt that emerging trends such as efficiency, productivity, sustainability and salt and fat reduction would be instrumental in successfully delivering their strategic objectives.⁵⁵ No such research could be found for South Africa.

There have been significant shifts in the foods consumed by South Africans, as packaged food and beverage innovation has accelerated since 1994.¹⁹ Globally, advances in science and technology regarding food processing have been made to meet consumer demand for minimally processed products and regulatory pressure to reduce the impact of food processing on the environment. Furthermore, advances in food manufacturing technologies and systems have also taken place to improve productivity, profitability and product quality. Significant capital investments have been made by the large and concentrated SAFBMI players in their local manufacturing facilities and supply chains. There are, however, no data to indicate in which technology areas these investments have been made and for what purpose. The objective of this study was therefore to understand how the SAFBMI (excluding alcohol) has invested and applied scientific and technological advances since 1994 to keep pace with the shifts in food consumption, consumer demands, competitive market dynamics and increasingly demanding regulations.

Material and methods

Food and beverage trade magazines were used as the source of scientific and technological advances that have been applied in the SAFBMI (excluding alcohol). These trade magazines were selected as they were the only relevant trade magazines available that spanned the time period under review, 1986 to 2012. They also report on activities in food and beverage processing and packaging technology in South Africa. *South African Food Review* (FR), a trade magazine for food and beverage manufacturers, informs readers about the introduction of new ingredients, equipment, products and packs. This trade magazine, published since 1986, provides objective commentary on local and international market trends and issues as well as relevant processing and packaging technologies.⁵⁶ *Food & Beverage Reporter* (F&BR), a

trade magazine focusing on strategy and technology for the industry, is the widest-circulating printed magazine for the food, beverage and packaging industries in southern Africa. It covers all sub-sectors in value-adding of food, such as milling, baking and processing, and has been published since 1995.⁵⁷ Even though these trade magazines are not peer-reviewed or scientifically validated on the application of science and technology by the SAFBMI, they do represent the common knowledge available over this time period and hence formed the basis for this study.

A database was built by capturing articles written from 1986 to 2012 for FR and from 1995 to 2012 for F&BR. Criteria for the selection of an article was that only feature articles (articles listed in the contents page) related to the application of science and technology by the SAFBMI, whose operations were in South Africa, were selected. The classification of the articles (referred to in the results and discussion section as 'activities') are an interpretation by the researcher based on the information shared in the article. The database was categorised according to the following: year, journal, volume, number, page, title, author, abstract, region, sector, company and activity type.

Activity type was classified as manufacturing, innovation or regulatory. A manufacturing activity was classified as an upgrade, new plant or new equipment. Innovation was split into the consumer driver behind the innovation and then into primary or secondary drivers. Regulatory activities were also classified and included legal requirements required by the South African government regarding the manufacture and sale of foodstuffs. Figure 2 further illustrates the categories of manufacturing activities. These were related to the rationale for the investment and type of process in which the activity was invested. Table 1 provides definitions and examples to explain the categories of activities. Graphs, charts and tables were then developed from the database using Microsoft Office Excel Pivots.

The data were analysed to determine trends in the application of science and technology. This study therefore provides an indicative, subjective review of the application of science and technology in the SAFBMI over the last three to four decades. It is not a scientifically validated or quantitative analysis, as trade magazines were used as the data source and not all applications of science and technology in the SAFBMI would have been published in trade magazines or as feature articles.

Results and discussion

The database contained over 370 entries for both FR and F&BR. There has been a significant increase in activities in the SAFBMI since 1986 (Figure 3), including activities related to both food and beverage production facilities, e.g. upgrades, new equipment or new plants and product and/or packaging innovations. There was a marked shift in the number of activities in the late 1990s and early 2000s as South Africa emerged as a new democracy. This shift is highly correlated with gross domestic product per capita over this time period (Figure 3).⁵⁸ Gross domestic product at purchasers' prices is the sum of gross value added by all resident producers in the economy, including any product taxes minus any subsidies not included in the value of the products.⁵⁸

In terms of activity split over the assessment period, innovation activities (54%) were more prevalent than production activities (46%). This difference could be for a number of reasons, including the focus areas in which the selected trade magazines were publishing in that period. Upgrades and new plants were the majority of advances as they relate to production facilities rather than new equipment.

Advances in science and technology

Capacity was the primary rationale for making advances to production facilities for F&BR data, with equal percentages for capacity, efficiency/productivity and quality for FR data (Figure 4). This finding would be expected for a growing food and beverage industry such as that of South Africa over the time period measured. For instance, the South African packaged food sector grew in volume by 15% in 5 years from 2007 to 2012.¹³

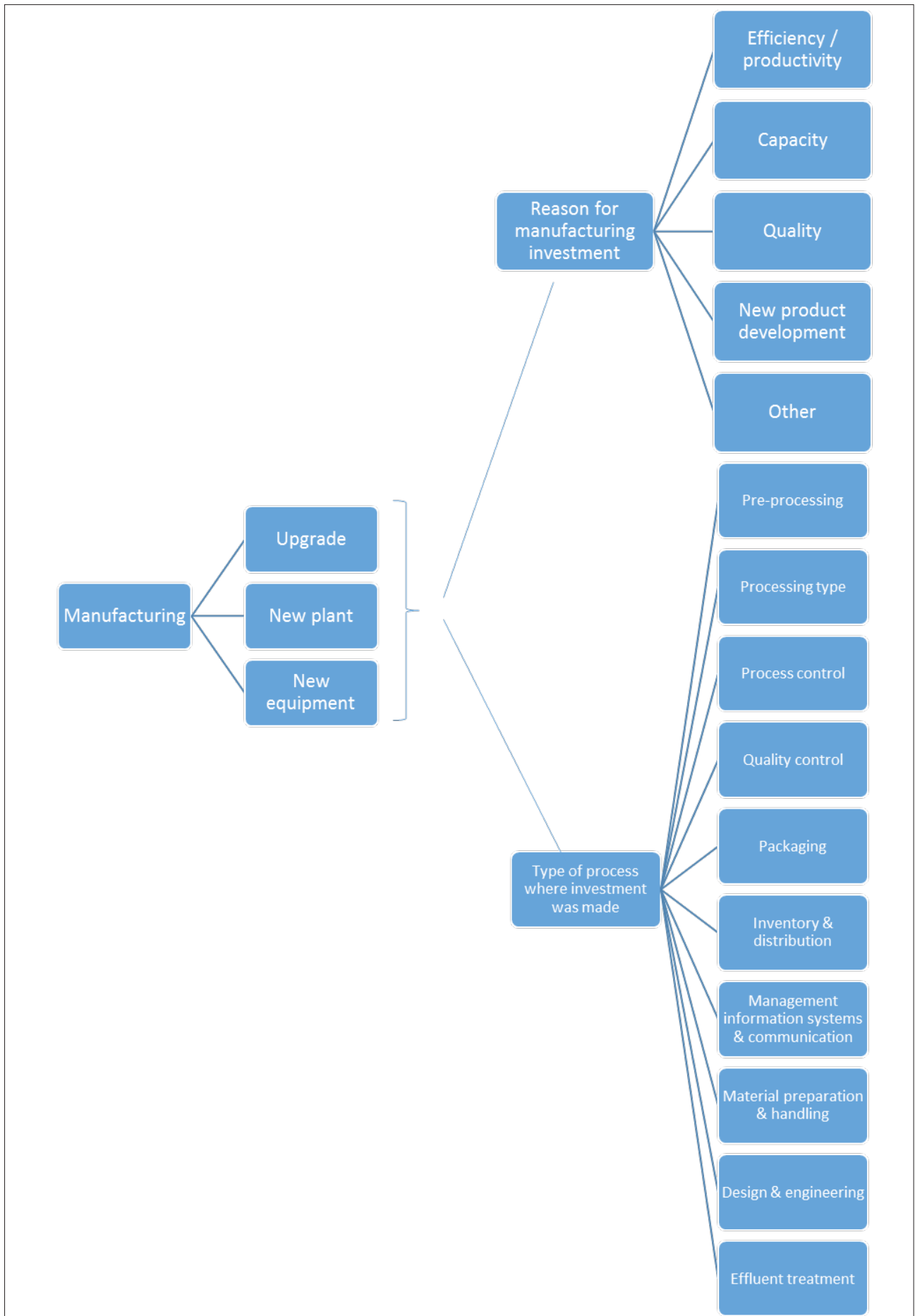


Figure 2: Further classification of manufacturing-related activities in the South African food and beverage manufacturing industry.

Table 1: Definitions and examples of the classification of activities in the South African food and beverage manufacturing industry (adapted and extended from Fryer & Versteeg³⁰ and Baldwin et al.⁴⁶) and general definitions

Term	Definition
Manufacturing	The process of converting raw materials and packaging into finished goods. It includes the classification of upgrade, new plant and new equipment for the purposes of efficiency/productivity, capacity, quality or other reasons.
Upgrade	Investment to improve existing facilities.
New plant	New building and manufacturing facilities.
New equipment	New food and beverage processing equipment.
Efficiency/productivity	Investments and upgrades to improve efficiency or increase productivity within the manufacturing facility.
Capacity	Investments and upgrades to increase yield or output within the manufacturing facility.
Quality	Investments and upgrades to improve product quality, e.g. reduction of defects, and to improve consistency or uniformity of product.
Other	Activities that did not fit into the criteria of efficiency/productivity, capacity or quality, e.g. replacing a building as a result of fire destruction, an innovation centre, developing new capabilities and sustainability.
Pre-processing	Raw material quality assessments and enhancement before processing, e.g. animal stress reduction, bran removal before milling wheat, electronic or ultrasonic grading, near-infrared analysis, colour assessment or sorting, electromechanical defect sorting and rapid testing techniques.
Processing type	Food processing is the procedure that food undergoes to go from its raw state to a finished product, ready for storage and shipping.
Thermal	Thermal processing is defined as the combination of temperature and time required to eliminate the desired number of microorganisms from a food product, ³³ e.g. aseptic processing, retortable, flexible pouches, infrared heating, ohmic heating and microwave heating.
Non-thermal	Food is treated at ambient or refrigeration temperatures and heat generation during the process is not substantial, ²⁷ e.g. chemical antimicrobials, ultrasonic techniques, high pressure (cold pasteurisation), deep chilling and irradiation.
Separation, concentration and water removal	The removal or concentration of water for the purposes of preserving a product for an extended period, e.g. membrane, filter technologies, centrifugation, ion exchange, vacuum microwave drying, water activity control and chromatographic separations.
Additives and ingredients	Examples include bio-ingredients and microbial cells.
Other	Specific food-processing technologies, e.g. drying, baking, milling, extrusion and coating.
Process control	Examples include automated sensor-based equipment for inspection/testing, automated statistical process control, machine vision, barcoding, programmable logic controllers and computerised process control.
Quality control	Observation techniques and activities used to fulfil requirements for quality, including process testing, laboratory testing and simulation, e.g. chromatography, monoclonal antibodies, DNA probes, rapid testing techniques and mathematical modelling of quality or safety.
Packaging	Various materials used to wrap or protect goods, including packaging assembly equipment.
Advanced materials	Examples include laminates, active packaging and multi-layer packaging.
Equipment	Examples include non-integrated or integrated electronically controlled machinery.
Glass	Examples include glass bottles of different shapes and sizes.
Metal	Examples include aluminium and tinplate cans.
Plastic	Examples include PET, HDPE, LDPE and PP.
Preservation	Examples include modified atmosphere.
Other	Mainly related to packaging artwork and design changes, e.g. graphics updates, limited edition packaging, label embossing or decoration.
Inventory and distribution	Inventory is related to the management of raw materials, work-in-process and finished goods. Distribution is the storage and delivery of products to the market place, e.g. barcoding and automated product handling.
Management information systems and communication	Examples include local area network, wide area network, inter-company computer networks, Internet marketing and promotion and other Internet-related networks.
Material preparation and handling	Examples include integrated electronically controlled machinery, individual electronically controlled non-integrated machinery and electronic detection of machinery failure.
Design and engineering	Examples include computer-aided design, computer-aided engineering, computer-aided manufacturing, computer-aided simulation and prototypes and digital representation of computer-aided design output used in procurement.
Effluent treatment	The treatment of waste water or air produced during the manufacturing process in order to reduce its environmental impact.
Innovation	Involves new product or packaging and includes renovation or re-development of an existing range.
Consumer driver – primary	The main rationale or proposition behind a new product or packaging development from a consumer perspective. Drivers were grouped into health and wellness, convenience, affordability, consumer preference, environment, food safety, product quality and urbanisation.
Consumer driver – secondary	The supporting or secondary rationale or proposition behind a new product or packaging development from a consumer perspective. Drivers were grouped into health and wellness, convenience, affordability, consumer preference, environment, food safety, product quality and urbanisation.
Regulatory	Related to all legal requirements of the South African government regarding the manufacture and sale of foodstuffs.
General definitions	
Soft drinks	Cola carbonates, non-cola carbonates, fruit/vegetable juices, bottled water, concentrates, ready-to-drink teas and sports and energy drinks.
Dairy	Drinking milk products, cheese, yoghurt and sour milk and baby milk formula.
Bakery	Baked goods (industrial and artisanal), biscuits and breakfast cereals.

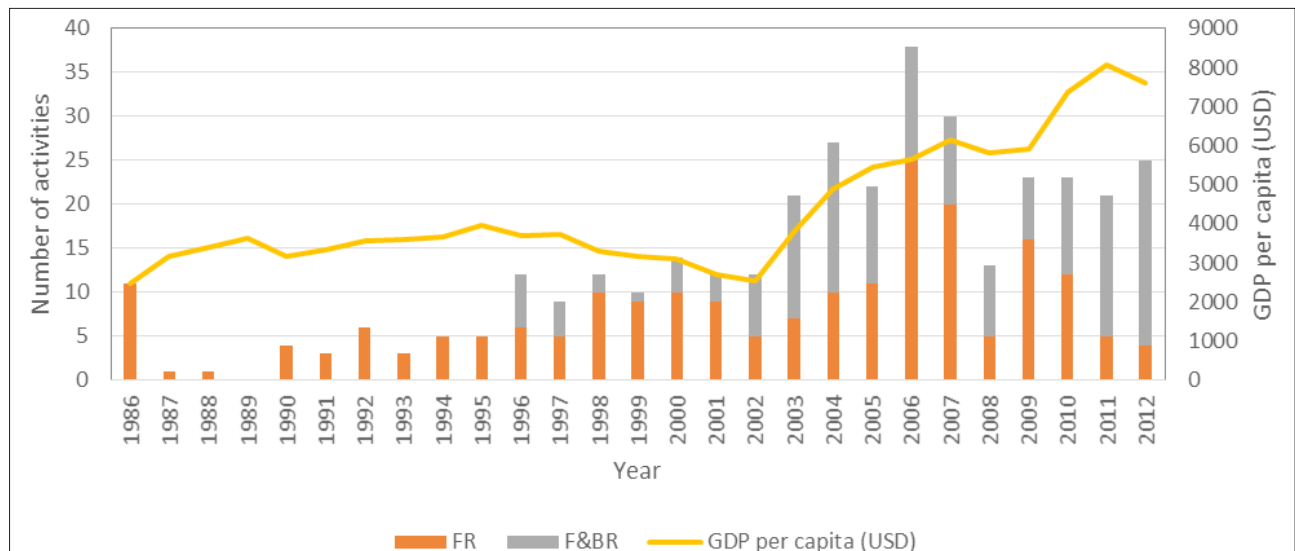
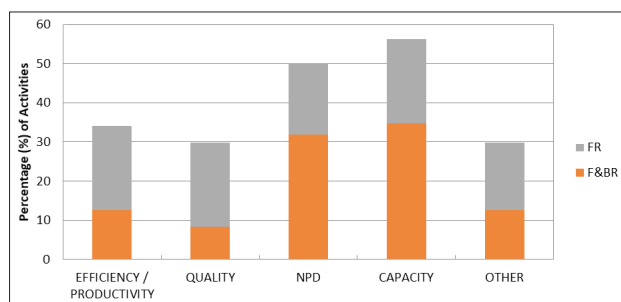


Figure 3: Number of activities recorded against gross domestic product (GDP) per capita obtained from *South African Food Review* (FR) from 1986 and *Food & Beverage Reporter* (F&BR) from 1996.



NPD, new product development

Figure 4: Rationale for production advances from *Food & Beverage Reporter* (F&BR) and *South African Food Review* (FR) data.

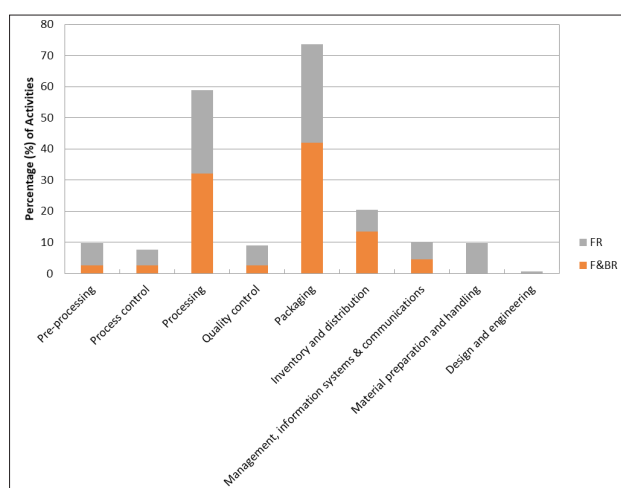


Figure 5: Focus areas of scientific and technological advances from *Food & Beverage Reporter* (F&BR) and *South African Food Review* (FR) data.

Packaging and processing were the main focus areas for scientific and technological advances and were consistent in both sets of data (Figure 5). Inventory and distribution and material preparation and handling were the second focus areas for the F&BR and FR data, respectively (Figure 5). Relating to inventory and distribution, examples such as implementation of warehouse management systems, refrigeration and cold store upgrades and new, larger, more sustainable

distribution centres were reported. Examples of material preparation and handling included automation, de-palletising and palletising equipment, conveying and mobile racking. Continued investment and advancement in science and technology is important to ensure products are safe and is necessary to continuously improve food safety management protocols. Examples of these food safety control measures noted in the study were detection and rejection of foreign objects through installation of metal detectors and colour/optical sorters, improved hygiene practices with the inclusion of cleaning-in-place and improved quality control practices with construction of in-house laboratories for product analysis and testing.

With a focus on scientific and technological advances in the processing arena, both FR and F&BR data indicated that the top three investment areas were (1) thermal, (2) other and (3) separation, concentration and water removal (Figure 6). This result is not surprising as thermal processing is the most extensively researched and widely utilised processing technology. Advances in thermal processing were mainly by the dairy industry, which was consistent in both FR and F&BR data. An example of advancement in thermal preservation is extended shelf life applied in the dairy industry since 2005. Extended shelf life involves fresh milk that has been treated in a specific manner, mostly by rapid heat/cool steam-infusion and mechanically via microfiltration or centrifugal separation, to reduce microbial count and pasteurise after being packaged under super-hygienic conditions.⁵⁹ The infusion heating process is designed to kill heat-resistant psychotropic aerobic spores, yielding a milk taste similar to that of pasteurised milk, with an extended shelf life under refrigerated conditions.⁵⁹ Other processing activities were mainly related to the milling and baking industry with improved milling, baking, coating technology, breakfast cereals and cereal bars. Finally, for separation, concentration and water removal activities, FR data revealed that the dairy sector had the most activity in this type of processing, with F&BR data indicating the dried processed food, fats and oils, and soft drinks (bottled water) sectors. An example is the use of osmotic concentration in the dried fruit industry to produce intermediate moisture foods that have a texture in-between that of dried and glacé fruit. An example of technology advances in the area of additives and ingredients is the application of advanced enzyme technology that eliminates the need for a cold chain for part-baked products and allows for smaller, more frequent bakes that last longer as a result of moisture retention. No feature articles related to the research of novel (i.e. not commercially available) thermal or non-thermal processing applications by the SAFBMI were noted over the time period documented. This finding could be for a number of reasons, such as the particular focus of the trade magazine, intellectual property protection and inability to disclose this information, insufficient numbers of research institutes (both academic and governmental) conducting research on novel processing or lack of funding and/or technical capability available in South Africa.

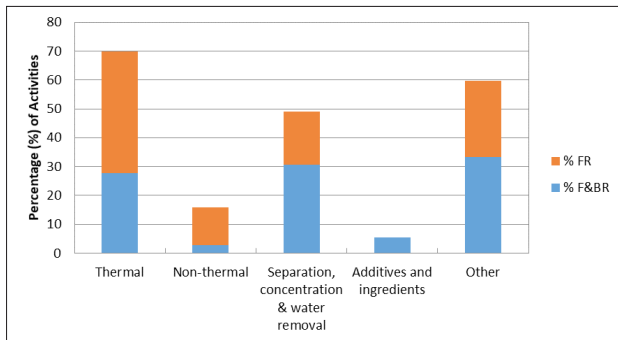


Figure 6: Scientific and technological advances by processing types from *Food & Beverage Reporter* (F&BR) and *South African Food Review* (FR) data.

From a packaging perspective, equipment, advanced materials and plastic were the main scientific and technological advances (Figure 7). Even though paper, glass, plastic and metals are South Africa's most utilised materials by volume,⁶⁰ the evidence suggested that there has been much innovation and application of advanced materials over the last three to four decades. In terms of advanced materials, FR data indicated that the dairy and meat sectors were active in this area, while F&BR data indicated the soft drink sector. A similar trend has been observed globally, with a shift away from rigid packaging to flexible packaging because of the benefits of being lightweight, improved barrier properties, ease of decoration, packaging format (variety and size) and dispensing options.⁶¹ Furthermore, the innovative flavour-release packaging, Sipahh®89, which enables flavour 'beads' to dissolve in the beverage as the liquid passes through the straw, was launched in South Africa over this time period. There was one report indicating the application of modified atmosphere packaging utilised in the meat sector. In terms of equipment, the soft drink sector, in both sets of data, was most prolific. High-speed bottling lines, flow wrappers and multi-head weighers were some of the investments made in the equipment area. Soft drinks and dairy were the sectors most active in the plastic packaging category in both sets of data. No examples of the application of smart packaging were recorded.

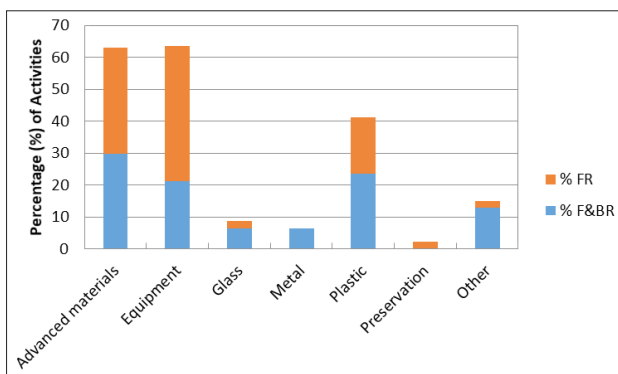


Figure 7: Scientific and technological advances by packaging types from *Food & Beverage Reporter* (F&BR) and *South African Food Review* (FR) data.

Consumer drivers behind innovation activities

Consumer drivers were derived from the rationale for product and/or packaging innovation. Both sets of data, across primary and secondary consumer drivers, indicated that the top three drivers of innovation were health and wellness, convenience and consumer preference (Figure 8). This finding was consistent with data from the Innova database (filtered for South Africa data), where convenience, health and choice were indicated as the main drivers.¹⁶

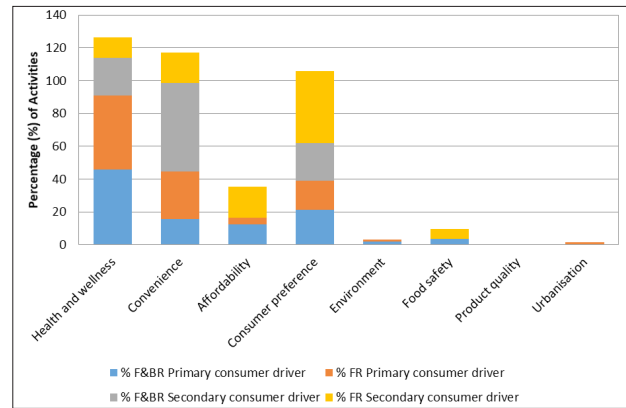


Figure 8: Primary and secondary consumer drivers for innovation from *Food & Beverage Reporter* (F&BR) and *South African Food Review* (FR) data.

Food consumption changes and advances in science and technology

In order to understand the link between food consumption changes in South Africa since 1994 and the application of science and technology and innovation activities, Figure 9 was generated. It is a combined data set of food consumption shifts taken from the Food and Agriculture Organization of the United Nations Statistics Division Food Balance Sheets and Euromonitor Passport from 1994/1999 to 2009/2012,¹⁹ plotted against the total number of activities (both production and innovation) by the SAFBML. Soft drinks and dairy have been the most active sectors in terms of advances in science and technology and innovation since 1986 (Figure 9). It was also evident that there has been a large shift in the consumption of soft drinks over a similar period of time (Figure 9). In terms of data captured in the database for the soft drinks sector, innovation activities were nearly double that of production. Demand for healthier products such as bottled water, juice nectars, ready-to-drink teas, low-calorie colas and non-cola carbonates have driven much of this innovation activity.⁶² Production activities implemented were related to upgrades, new plants and equipment to increase capacity and productivity. The soft drink category grew in volume by 14.9% in 5 years, from 4130 million litres in 2007 to 4745 million litres in 2012.⁶² To keep pace with that level of volume growth, it would make sense that upgrades were made to increase capacity and productivity.

Even though the dairy sector experienced a large increase in activities related to advances in science and technology and innovation, it did not experience large increases in dairy consumption overall. However, yoghurt and sour milk, which contributed a small percentage to the dairy category, experienced large increases in consumption (74%).¹⁹ In terms of dairy, there was a balanced number of innovation- and production-related activities recorded in the database, with a focus on upgrades, new plants and equipment. The main reasons for the investment were related to increasing capabilities for new product development followed by productivity and capacity. The growing consumer demand for health and convenience drove new product developments in dairy over this time period.⁶³ Spoonable yoghurt experienced growth of up to 10% in 2012⁶⁴ and the majority of innovation activities recorded in this study were related to yoghurt innovation.

Bakery was the third-most active sector with overall consumption increases of 6.4% because of declines in consumption of packaged/industrial bread as consumers traded up to ready-to-eat breakfast cereals or down to maize meal, depending on their income situation (Figure 9). However, sub-categories within the bakery sector, such as biscuits and breakfast cereals, saw large increases in consumption of 57% and 36%, respectively, albeit off a small consumption base.¹⁹ Furthermore, there was a balanced number of innovation- and production-related activities recorded in the database, with a focus mainly on upgrades and new plants. The main reasons for investment were related to capacity and quality improvements. Baked goods and biscuits increased in volume by 18.5% and 38.5%, respectively, in 5 years, from 2007 to 2012, and hence the need for capacity investments.^{65,66}

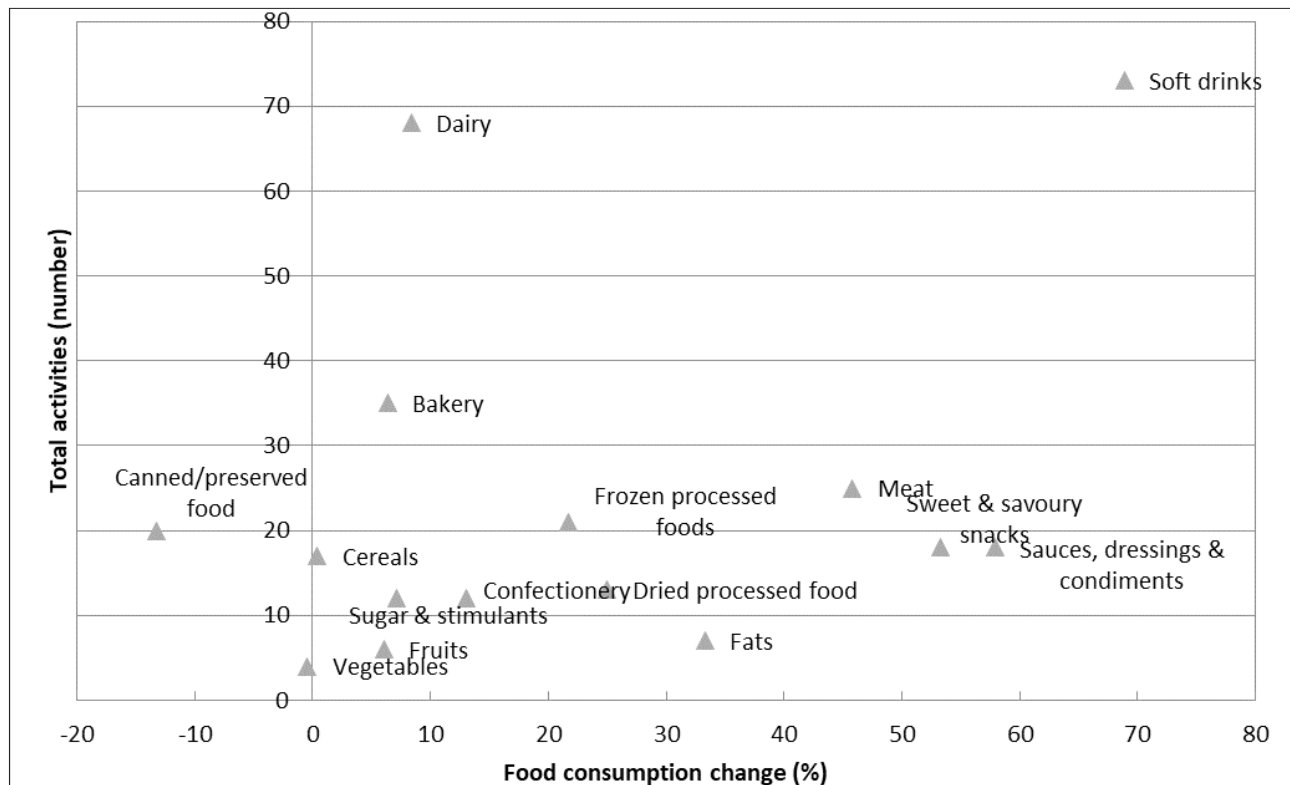


Figure 9: Number of activities by sector combined for both *Food & Beverage Reporter* and *South African Food Review* data since 1986 compared with food consumption shifts from 1994/1999 to 2009/2012.

Conclusion

There have been dramatic and significant shifts in food consumption in South Africa since 1994, especially in the food categories soft drinks; sauces, dressings and condiments; sweet and savoury snacks; meat; and fats and oils. The SAFBMI responded by:

- Upgrading manufacturing facilities and building new plants to increase capacity, deliver new products and improve efficiencies and product quality and safety.
- Investing in processing, especially in the areas of thermal processing and other processing activities such as baking, milling and separation, concentration and water removal.
- Investing in the areas of packaging, including mainly equipment, advanced materials such as laminates and multilayer packaging and plastics.
- Innovating through new product and packaging launches to meet consumer drivers of health and nutrition, convenience and changing consumer preferences.

The packaged food and beverage industry became active relating to facility upgrades and building of new plants after 1994, with a sharp increase in product launches observed in the late 2000s.¹⁶ The dairy, soft drinks and bakery sectors have been the most active. As far as advances in food preservation are concerned, there is evidence of the application of commercially available new preservation technologies such as thermal processing to extend the shelf life of milk and modified atmosphere packaging. Even though there have been advances in science and technology globally relating to novel processing, preservation and packaging, we found a low level of experimentation of these non-commercially available advances by the SAFBMI. From this research, South Africa would appear to be keeping pace with the advances in food manufacturing in the areas of automation, process and quality and food safety management, material handling as well as in the establishment of

larger distribution centres with warehouse management systems, with the overall objective of improving production efficiencies.

It is important to recognise the limitation of this study, which is that it is based on common knowledge in the area of science and technology advances by SAFBMI available from trade magazines. We did not assess technical competence nor culture, which is fundamental to delivering innovative, safe and consistent quality products, efficiently.

In the future, investment and research in food science and technology by South African food and beverage manufacturers, research institutes and universities will continue to be necessary to achieve current and impending regulatory requirements as well as increasing consumer demand for packaged foods and beverages. Adoption of novel processing methods could yield product quality, safety, nutrition, productivity and capacity benefits.

Acknowledgements

We thank Johan Visser and Anee Sieberhagen at Nampak Research & Development for providing access to their extensive library; Lu Ann Williams from Innova Market Insights for access to the Innova database; and Mars Africa for providing time and financial assistance.

Authors' contributions

L-C.R-R. was the project leader and responsible for data collection, analysis and validation and writing the manuscript; G.O.S. was the student supervisor and made editorial contributions; L-C.R-R., G.O.S. and N.V. were responsible for project design (conceptualisation and methodology).

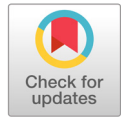
References

1. The Presidency, Republic of South Africa. Twenty year review: South Africa: 1994–2014 [document on the Internet]. Pretoria: The Presidency; 2014 [cited 2014 Oct 12]. Available from: <http://www.dpme.gov.za/news/Documents/20%20Year%20Review.pdf>

2. Reardon T, Gulati A. The supermarket revolution in developing countries policies for "Competitiveness with Inclusiveness". International Food Policy Research Institute (IFPRI) Policy Brief 2. East Lansing, MI: Michigan State University; 2008.
3. Shisana O, Labadarios D, Rehle T, Simbayi L, Zuma K, Dhansay A, et al. South African National Health and Nutrition Examination Survey (SANHANES-1). Cape Town: HSRC Press; 2013.
4. Kennedy G, Nantel G, Shetty P. Globalisation of food systems in developing countries: A synthesis of country case studies. Food and Nutrition Paper 83. Rome: Food and Agriculture Organization of the United Nations; 2004.
5. Chopra M, Galbraith S, Darnton-Hill I. A global response to a global problem: The epidemic of over nutrition. Bull World Health Organ. 2002;80(12):952–958. <http://dx.doi.org/10.1590/S0042-96862002001200009>
6. Lang T. Food industrialisation and food power: Implications for food governance. Gatekeeper Series No. 114. London: International Institute for Environment and Development, Natural Resources Group and Sustainable Agriculture and Rural Livelihoods Programme; 2004.
7. Gardner G, Halweil B. Underfed and overfed: The global epidemic of malnutrition. Worldwatch Paper 150. Washington DC: Worldwatch Institute; 2000.
8. Igumbor EU, Sanders D, Puoane TR, Tsolekile L, Schwarz C, Purdy C, et al. "Big food," the consumer food environment, health, and the policy response in South Africa. PLoS Med. 2012;9(7), e1001253, 7 pages. <https://doi.org/10.1371/journal.pmed.1001253>
9. Ntloedibe M. South Africa's food processing sector offers opportunities for U.S. exporters. Pretoria: USDA Foreign Agricultural Service Global Agricultural Information Network; 2011.
10. Alexander E, Yach D, Mensah GA. Major multinational food and beverage companies and informal sector contributions to global food consumption: Implications for nutrition policy. Glob Health. 2011;7, Art. #26, 8 pages. <http://dx.doi.org/10.1186/1744-8603-7-26>
11. Stuckler D, McKee M, Ebrahim S, Basu S. Manufacturing epidemics: The role of global producers in increased consumption of unhealthy commodities including processed foods, alcohol and tobacco. PLoS Med. 2012;9(6), e1001235, 8 pages. <http://dx.doi.org/10.1371/journal.pmed.1001235>
12. Mather C. The growth challenges of small and medium enterprises (SMEs) in South Africa's food processing complex. Development Southern Africa. 2005;22(5):607–622. <http://dx.doi.org/10.1080/03768350500364208>
13. Euromonitor Passport packaged food in South Africa [database on the Internet]. c2012 [cited 2014 Oct 12]. Available from: <http://www.portal.euromonitor.com/Portal/Default.aspx>
14. Statistics South Africa. Large sample survey of the manufacturing industry 2001. Pretoria: Statistics South Africa; 2001.
15. Statistics South Africa. Manufacturing industry: Financial, 2008. Pretoria: Statistics South Africa; 2010.
16. Innova. Innova database search [database on the Internet]. 2014 [cited 2014 Jan 23]. Available from: <http://new.innovadatabase.com/search/searchresults.rails>
17. Sonneveld K. What drives (food) packaging innovation? Packag Technol Sci. 2000;13:29–35. [http://dx.doi.org/10.1002/\(SICI\)1099-1522\(200001/02\)13:1<29::AID-PTS489>3.0.CO;2-R](http://dx.doi.org/10.1002/(SICI)1099-1522(200001/02)13:1<29::AID-PTS489>3.0.CO;2-R)
18. Andrews J. Technology: The core ingredient in 'natural' foods [webpage on the Internet]. Cape Town: FOODSuff South Africa; 2016 [updated 2016 January 21; cited 2016 January 29]. Available from: <http://www.foodstuffs.co.za/news-stuff/food-science-and-technology-stuff/4363-technology-the-core-ingredient-in-natural-foods>
19. Ronquest-Ross L-C, Vink N, Sigge GO. Food consumption changes in South Africa since 1994. S Afr J Sci. 2015;111(9/10), Art. #2014-0354, 12 pages. <https://doi.org/10.17159/sajs.2015/20140354>
20. World Wildlife Fund. Agriculture: Facts and trends South Africa [document on the Internet]. c2012 [cited 2012 Dec 21]. Available from: http://awsassets.wwf.org.za/downloads/facts_brochure_mockup_04_b.pdf
21. Bureau for Food and Agricultural Policy (BFAP). The South African agricultural baseline [homepage on the Internet]. c2013 [cited 2013 Nov 04]. Available from: http://www.bfap.co.za/documents/baselines/BFAP_Baseline_2013
22. Kruger HS, Puoane T, Senekal M, Van der Merwe MT. Obesity in South Africa: Challenges for government and health professionals. Public Health Nutr. 2005;8(5):491–500. <http://dx.doi.org/10.1079/PHN2005785>
23. Popkin BM. Global nutrition dynamics: The world is shifting rapidly toward a diet linked with noncommunicable disease. Am J Clin Nutr. 2006;84(2):289–298. <https://doi.org/10.1093/ajcn/84.2.289>
24. Shisana O, Labadarios D, Rehle T, Simbayi L, Zuma K, Dhansay A, et al. South African National Health and Nutrition Examination Survey (SANHANES-1). Cape Town: HSRC Press; 2013.
25. Toepfl S, Mathys A, Heinz V, Knorr D. Review: Potential of high hydrostatic pressure and pulsed electric fields for energy efficient and environmentally friendly food processing. Food Rev Int. 2006;22(4):405–423. <http://dx.doi.org/10.1080/87559120600865164>
26. Lado BH, Yousef AE. Alternative food-preservation technologies: Efficacy and mechanisms. Microbes Infect. 2002;4(4):433–440. [http://dx.doi.org/10.1016/S1276-4579\(02\)01557-5](http://dx.doi.org/10.1016/S1276-4579(02)01557-5)
27. Rodrigo D, Sampedro F, Silva A, Palop A, Martinez A. New food processing technologies as a paradigm of safety and quality. Brit Food J. 2010;112(5):467–475. <http://dx.doi.org/10.1108/00070701011043727>
28. Barbosa-Cánovas GV, Juliano P. Adaptation of classical processes to new technical developments and quality requirements. J Food Sci. 2004;69(5):240–250. <http://dx.doi.org/10.1111/j.1365-2621.2004.tb10715.x>
29. Heinz V, Alvarez I, Angersbacha A, Knorr D. Preservation of liquid foods by high intensity pulsed electric fields – basic concepts for process design. Trends Food Sci Tech. 2002;12(3–4):103–111. [http://dx.doi.org/10.1016/S0924-2244\(01\)00064-4](http://dx.doi.org/10.1016/S0924-2244(01)00064-4)
30. Fryer PJ, Versteeg C. Processing technology innovation in the food industry. Innov Manag Policy P. 2008;10(1):74–90. <https://doi.org/10.5172/imp.453.10.1.74>
31. Floros JD, Newsome R, Fisher W, Barbosa-Cánovas GV, Chen H, Dunne CP, et al. Feeding the world today and tomorrow: The importance of food science and technology, an IFT scientific review. Compr Rev Food Sci F. 2010;9(5):572–599. <http://dx.doi.org/10.1111/j.1541-4337.2010.00127.x>
32. Tucker G, Featherstone S. Essentials of thermal processing. Chichester: Wiley-Blackwell; 2011.
33. Zhang Q, Barbosa-Cánovas GV, Swanson BG. Engineering aspects of pulsed electric field pasteurisation. J Food Eng. 1995;25(2):261–281. [http://dx.doi.org/10.1016/0260-8774\(94\)00030-D](http://dx.doi.org/10.1016/0260-8774(94)00030-D)
34. Knorr D. New developments in industrial processing. Brussels: European Parliament Science and Technology Options Assessment (STOA); 2008. p. 23–26.
35. Koutchma T. Advances in ultraviolet light technology for non-thermal processing of liquid foods. Food Bioprocess Tech. 2009;2(2):138–155. <http://dx.doi.org/10.1007/s11947-008-0178-3>
36. Gayán E, Álvarez I, Condón S. Inactivation of bacterial spores by UV-C light. Innov Food Sci Emerg. 2013;19:140–145. <http://dx.doi.org/10.1016/j.ifset.2013.04.007>
37. Ansari IA, Datta K. An overview of sterilisation methods for packaging materials used in aseptic packaging systems. Food Bioprod Process. 2003;81(1):57–65. <http://dx.doi.org/10.1205/096030803765208670>
38. Brody AL, Bugusu B, Han JH, Sand CK, McHugh TH. Scientific status summary. J Food Sci. 2008;73(8):R107–R116. <http://dx.doi.org/10.1111/j.1750-3841.2008.00933.x>
39. Martínez-Romero D, Guillén F, Castillo S, Valero D, Serrano M. Modified atmosphere packaging maintains quality of table grapes. J Food Sci. 2003;68(5):1838–1843. <http://dx.doi.org/10.1111/j.1365-2621.2003.tb12339.x>
40. Mahalik NP. Advances in packaging methods, processes and systems. Challenges. 2014;5(2):374–389. <http://dx.doi.org/10.3390/challe5020374>
41. Rijk R. Legislative issues relating to smart packaging. In: Kerry J, Butler P, editors. Smart packaging technologies for fast moving consumer goods. Chichester: John Wiley & Sons; 2008. p. 305–320. <https://doi.org/10.1002/9780470753699.ch17>


42. Sozer N, Kokini JL. Nanotechnology and its applications in the food sector. *Trends Biotechnol.* 2008;27(2):82–89. <http://dx.doi.org/10.1016/j.tibtech.2008.10.010>
43. Meetoo DD. Nanotechnology and the food sector: From the farm to the table. *Emir J Food Agric.* 2011;23(5):387–403.
44. O'Sullivan MG, Kerry JP. Smart packaging technologies for beverage products. In: Kerry J, Butler P, editors. *Smart packaging technologies for fast moving consumer goods*. Chichester: John Wiley & Sons; 2008. p. 211–229. <https://doi.org/10.1002/9780470753699.ch12>
45. Steeman A. The self-cooling technology and the future – part 3 [webpage on the Internet]. c2012 [updated 2012 Dec 17; cited 2016 Jun 16]. Available from: <https://bestinpackaging.com/2012/12/17/the-self-cooling-technology-and-the-future-part-3/>
46. Marsh K, Bugusu B. Food packaging – roles, materials, and environmental issues. *J Food Sci.* 2007;72(3):R39–R55. <http://dx.doi.org/10.1111/j.1750-3841.2007.00301.x>
47. Sterling S. *Field guide to sustainable packaging*. Chicago, IL: Summit; 2007.
48. Baldwin J, Sabourin D, West D. *Advanced technology in the Canadian food processing industry*. Statistics Canada catalogue no. 88-518-XIE. Ottawa: Agriculture and Agri-Food Canada; 1999.
49. Sani MS, Aziz FA. Advanced manufacturing systems in food processing and packaging industry. *Mater Sci Eng.* 2013;46, Art. #012042, 9 pages. <http://dx.doi.org/10.1088/1757-899X/46/1/012042>
50. Fellows PJ. *Food processing technology principles and practice*. 3rd ed. Cambridge: Woodhead; 2009.
51. Leisemann Immel R. Programmable logic controller evolution [webpage on the Internet]. c2012 [updated 2012 Mar 19; cited 2016 Jun 11]. Available from: <http://www.mbtmag.com/article/2012/03/programmable-logic-controller-evolution>
52. Alpern P. Food for thought in robot technology. *Industry Week.* 2010 May 17;6:259.
53. Melendez K. Retail logistics: The advantages of a central distribution center [webpage on the Internet]. c2013 [updated 2013 Jul 09; cited 2016 Jul 03]. Available from: <http://info.mkmdistribution.com/blog/bid/314977/Retail-Logistics-The-advantages-of-a-central-distribution-center>
54. McGovern CM, Manley M. Classification of maize kernel hardness using near infrared hyperspectral imaging. *J Near Infrared Spectrosc.* 2012;20(5):529–535. <http://dx.doi.org/10.1255/jnirs.1018>
55. Leatherhead Food RA. Survey: What's most important when it comes to investing in new technology [webpage on the Internet]. Cape Town: FOODStuff South Africa; 2012 [updated 2012 Sep 20; cited 2012 Oct 02]. Available from: <http://foodstuffsa.co.za/news-stuff/food-science-and-technology-stuff/2385-survey-whats-most-important-when-it-comes-to-investing-in-new-technology>
56. South African Food Review. The journal for food and beverage manufacturers [webpage on the Internet]. Cape Town: New Media Publishing; 2014 [cited 2014 Oct 06]. Available from: <http://foodreview.co.za/>
57. Food & Beverage Reporter [webpage on the Internet]. c2014 [cited 2014 Oct 06]. Available from: <http://www.fbreporter.co.za>
58. World Bank. South Africa data [webpage on the Internet]. c2015 [cited 2015 Nov 15]. Available from: <http://data.worldbank.org/country/south-africa>
59. Neall B. ESL milk debuts in SA. *South African Food Review.* 2005;32(10):24.
60. PackagingSA. Packaging industry stats [webpage on the Internet]. c2014 [cited 2014 Oct 12]. Available from: <http://www.pacsa.co.za/info-library/packaging-industry-stats/>
61. Global high barrier packaging film consumption forecast to reach 2.24 million tonnes by 2019 [webpage on the Internet]. c2014 [updated 2014 Apr 28; cited 2016 Feb 02]. Available from: <http://www.smitherspira.com/news/2014/april/high-barrier-packaging-film-consumption-forecast>
62. Euromonitor International. Soft drinks in South Africa [database on the Internet]. c2013 [cited 2014 Oct 12]. Available from: <http://www.portal.euromonitor.com/Portal/Default.aspx>
63. Euromonitor International. Drinking milk products in South Africa [database on the Internet]. c2013 [cited 2014 Oct 12]. Available from: <http://www.portal.euromonitor.com/Portal/Default.aspx>
64. Euromonitor International. Yoghurt and sour milk products in South Africa [database on the Internet]. c2013 [cited 2014 Oct 12]. Available from: <http://www.portal.euromonitor.com/Portal/Default.aspx>
65. Euromonitor International. Baked goods in South Africa [database on the Internet]. c2013 [cited 2014 Oct 12]. Available from: <http://www.portal.euromonitor.com/Portal/Default.aspx>
66. Euromonitor International. Biscuits in South Africa [database on the Internet]. c2013 [cited 2014 Oct 12]. Available from: <http://www.portal.euromonitor.com/Portal/Default.aspx>





First ground-based observations of sprites over southern Africa

AUTHORS:

Stanislaus Nnadih^{1,2} 

Mike Kosch^{3,4,5}

Peter Martinez²

Jozsef Bor⁶

AFFILIATIONS:

¹African Regional Centre for Space Science and Technology Education, Ile Ife, Nigeria

²SpaceLab, Department of Electrical Engineering, University of Cape Town, Cape Town, South Africa

³Space Science Division, South African National Space Agency, Hermanus, South Africa

⁴Department of Physics, University of Western Cape, Cape Town, South Africa

⁵Physics Department, Lancaster University, Lancaster, United Kingdom

⁶Research Centre for Astronomy and Earth Science, Hungarian Academy of Science, Budapest, Hungary

CORRESPONDENCE TO:

Stanislaus Nnadih

EMAIL:

Nndsta001@myuct.ac.za

DATES:

Received: 29 Aug. 2017

Revised: 19 Jan. 2018

Accepted: 06 June 2018

Published: 11 Sep. 2018

KEYWORDS:

lightning discharge; mesosphere; atmosphere; thunderstorm

HOW TO CITE:

Nnadih S, Kosch M, Martinez P, Bor J. First ground-based observations of sprites over southern Africa. *S Afr J Sci.* 2018;114(9/10), Art. #4272, 6 pages. <https://doi.org/10.17159/sajs.2018/4272>

ARTICLE INCLUDES:

✓ Supplementary material

× Data set

FUNDING:

National Research Foundation (South Africa)

© 2018. The Author(s).
Published under a Creative Commons Attribution Licence.

Sprites are the optical signatures of electrical discharges in the mesosphere triggered by large lightning strikes associated with thunderstorms. Since their discovery in the late 1980s, sprites have been observed extensively around the world, although very few observations of sprites from Africa have been documented in the literature. In this paper, we report the first ground-based recorded observations of sprites from South Africa. In 2 out of the 22 nights of observations (11 January and 2 February 2016), about 100 sprite elements were recorded from Sutherland in the Northern Cape, comprising different morphologies (carrot (55%), carrot/column (11%), unclassified (21%), column (13%). The sprites were triggered by positive cloud-to-ground lightning strikes, which had an average peak value of ~ 74 kA and were observed at distances from ~ 400 km to 800 km. The estimated charge moment change of the lightning discharges associated with these events was in agreement with the threshold for dielectric breakdown of the mesosphere and correlates well with the observed sprite brightness.

Significance:

- The first ground-based recording of sprite events over southern Africa.
- It is suggested that the intensity of the events is proportional to the lightning stroke current.

Introduction

Anecdotal accounts of sporadic brief optical emissions appearing above thunderstorms have been found in the scientific literature for many years.¹⁻⁴ Such phenomena were also noted by South African observers. For instance, in 1937, Malan⁵ reported seeing a different kind of lightning in Johannesburg, at an approximate altitude of 50 km. He noted:

On the evening of 2 January 1937, there was a long cloud bank about 100 km to the northeast... During the span of an hour and a half, a long and weak streamer of reddish hue appeared about ten times in the upper cloud simultaneously with a weak illumination of the same cloud. I assume the weak illumination was due to cloud-to-ground discharge (CG) behind the hills.

These anecdotal pieces of evidence were not corroborated until 1989, when Prof. J. Winckler's group accidentally discovered a flash of light, illuminating the sky above a distant thunderstorm.⁶ This discovery sparked interest in such phenomena and their potential dangers to crewed spacecraft during take-off or landing. This interest later led to the discovery of all other optical phenomena in the stratosphere and mesosphere (ELVES, blue jets, halos, and gigantic jets), which were collectively termed transient luminous events.^{7,8}

Sprites are short-lived optical phenomena occurring above large thunderstorms. They are often described as large electrical discharges in the mesosphere.⁹ When observed with a high-speed camera,^{10,11} they appear to move downwards (for positive streamers) and then upwards (for negative streamers) with a speed in excess of 10^7 m/s. They are mostly associated with positive cloud-to-ground (+CG) lightning strikes, which transfer a positive charge to the ground¹²⁻¹⁴, although some studies have revealed that, on rare occasions, negative cloud-to-ground (-CG) lightning strikes can also initiate sprites^{15,16}.

Sprite research has been actively conducted in the USA, Asia, Europe, South America and Australia.^{12,17-20} Sprites have also been successfully studied from the International Space Station over central Africa, northern Australia and South America, as well as over the Pacific and Indian Oceans.²¹⁻²³ These studies have led to many discoveries (e.g. the ELF radiation produced by the electrical current in sprites²⁴ and the energy deposited by large sprites is of the order 1 GJ/event²⁵), which have been documented in numerous scientific publications describing the sprite phenomenon and its dynamics²⁶⁻²⁸.

Some of these studies suggest that, apart from contributing to the global electric circuit, sprite initiation can also affect the chemistry of the middle atmosphere by altering the composition of the oxides of nitrogen and hydrogen. These chemical changes could have an impact on the heating or cooling of this region²⁹ or interfere with long-distance communication through the lower ionosphere³⁰. Thus, it is important to determine the occurrence rate and geographical distribution of sprites and the energy associated with these events in order to verify these assertions. The global occurrence rate for sprites has been estimated to be 1–3 sprite events every minute.^{25,31-33}

Despite being a lightning-rich continent³⁴, there is currently no active ongoing sprite-related research in Africa, although there were a few observations carried out in West Africa during the 2006 African Monsoon Multidisciplinary Analysis campaign³⁵. In this paper, we describe the observations that led to the first ground-based recording of sprites over southern Africa.

Observations

Instrumentation

The optical instrument used for the observations reported in this paper was a Watec 910Hx camera, which has a 40-ms time resolution and was operated with a 80-ms temporal resolution in order to improve the signal-to-noise level. This camera is a standard low-light commercially available monochromatic charge-coupled device (CCD) camera (Figure 1), which is commonly used by observers of transient luminous events in Europe, the USA and Israel.^{36,37} The CCD detector array of 352 by 288 pixels was oriented with the long axis of the detector in the horizontal plane. The camera was fitted with a 8.0 mm F/1.4 lens. The field of view on the detector was 35.3° horizontal and 26.6° vertical, as determined by reference to known background stars on the image. This gives a horizontal/vertical angular resolution of 0.10/0.09° per pixel. The camera records video frames with 8-bit resolution.



Figure 1: The camera used during the sprite observations.

The camera was connected through a video digitiser (Pinnacle Dazzle) to a computer that ran the Sonotaco sprite capture software version 2 (http://sonotaco.com/soft/e_index.html). The software has a triggering mechanism that allows it to read out the CCD continuously, but only stores data according to the detection threshold that was set by the operator. To ensure the accuracy of the timing system, the computer was connected to the Network Timing Protocol server of the South African Astronomical Observatory (SAAO) in Sutherland, which has a time resolution of approximately 100 μ s.

Camera pointing angle

Previous studies have shown that sprites occur within the mesosphere at an altitude range of ~40–90 km, with maximum brightness at an altitude of ~68 km lasting several milliseconds.^{38,39} Thus, it was necessary to point the camera at the right elevation for a successful observation. Equation 1 relates the pointing elevation angle for the camera to different distances and altitudes for a given viewing geometry as shown in Kosch et al.⁴⁰

$$\tan[90-\theta] = \left\{ (h + Re) \sin \left[\frac{d}{Re} \right] \right\} / \left\{ (h + Re) \cos \left[\frac{d}{Re} \right] - Re \right\}, \text{ Equation 1}$$

where θ is the observation elevation angle; h is the altitude above ground, $Re = 6370$ km and is the radius of the earth, and d is the distance of the storm from the camera along a great circle arc.

Thus, for $h = 100$ km (lower boundary of the nighttime ionosphere):

$$d = 100 \text{ km} \Rightarrow \theta = 44.3^\circ$$

$$d = 200 \text{ km} \Rightarrow \theta = 25.5^\circ$$

$$d = 300 \text{ km} \Rightarrow \theta = 17.0^\circ$$

$$d = 400 \text{ km} \Rightarrow \theta = 12.1^\circ$$

The orientation of the camera (azimuth and elevation angle) was determined during the analysis, by identifying the stars recorded in the image, which can be done to single-pixel accuracy.

Observation technique

This instrument was set up at the SAAO in Sutherland, Northern Cape, South Africa (20.8117 E, 32.3872 S) and operated during the seasonal peak lightning period in southern Africa from December 2015 to February 2016.⁴¹ SAAO was an ideal location for such observations because of its altitude (1798 m) above sea level. This location was also chosen because of its clear dark night sky that is free of light pollution from cities, which allows the optical instrument to observe any transient luminous events up to a 900 km radius from SAAO.

When to operate the camera and where to point it was determined in near real time by observing storm activities over southern Africa using Meteosat infrared imagery, which is made available in near real time by EUMETSAT (<http://www.eumetsat.int/website/home/index.html>). This information was also complemented by lightning forecast data from the SAT24 website (<http://en.sat24.com/en/forecastimages/afrika/forecastlightning>). The Meteosat infrared imagery is automatically updated every 15 min, making it possible to track a storm almost in real time. The infrared band was chosen to enable us to identify the brightest part of the cloud because the brighter the cloud, the colder and the more likely it is to produce lightning strokes.⁴²

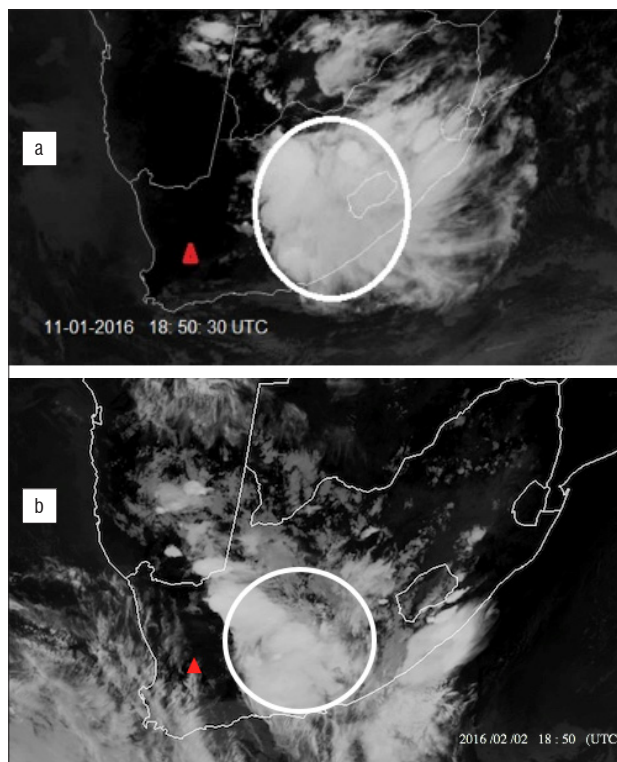


Figure 2: Meteosat infrared imagery showing the thunderstorms tracked on the nights of (a) 11 January 2016 and (b) 2 February 2016. The triangle shows the observation point. The circle shows the storm, which was ~350 km from the observation point in (a) and ~280 km in (b).

For each night of sprite observation, the Meteosat imagery over southern Africa was examined for storm activity to select the camera viewing azimuth. For thunderstorms beyond 400 km from the observation point, an elevation angle of 15° was used to avoid pointing the bottom edge of the camera's 26.6° vertical field of view below the horizon and imaging lightning strokes that cause false detections. As most of the storms observed during the campaign occurred beyond 250 km away, the value used for the elevation angle was set to ~15°, whereas the azimuth was regularly adjusted based on the nightly lightning strike forecasts. Figure 2 shows some examples of the Meteosat imagery used. As the storm weakened, moved or became more active in another area, the camera was manually adjusted to point in the direction of the most intense lightning activity. Observations of a particular storm typically lasted several hours. Depending on the number of cells that make up a cumulonimbus cloud, an average active storm can remain energetic up to 3 h before weakening (Lennard C 2016, oral communication, January 15).

Results

On 2 of the 22 nights of observations, a total of 54 video frames comprising ~100 sprites were recorded. By visual inspection, these events exhibit different shapes/structures, which include carrot (55%), column (13%), carrot/column (11%) and unclassified (21%). The naming of these groups depends on the final structure of the emission as described by Bor⁴³. Figure 3 shows an example of the different types

of sprites recorded during these observations, and Table 1 presents the structural statistics of all the events.

Table 1: Structural statistics of recorded events

Sprite type	Distribution (%)	Links to video frames of the events
Carrot	55	20160111_18:53:49.4
Carrot/column	11	20160111_19:17:09.8
Unclassified	21	20160111_19:49:19.7
Column	13	20160111_19:38:32.5

Data analysis

The sprite images were superimposed on the star catalogue of the Sonotaco sprite analyser software (http://sonotaco.com/soft/e_index.html) version 2. This software uses the star coordinate data that were extracted from the SKY2000 Master Catalogue, version 4. The stars on the sprite images were manually aligned to the stars in this catalogue to within single-pixel resolution to estimate the azimuth, elevation and location (longitude and latitude) of each of these sprite elements.

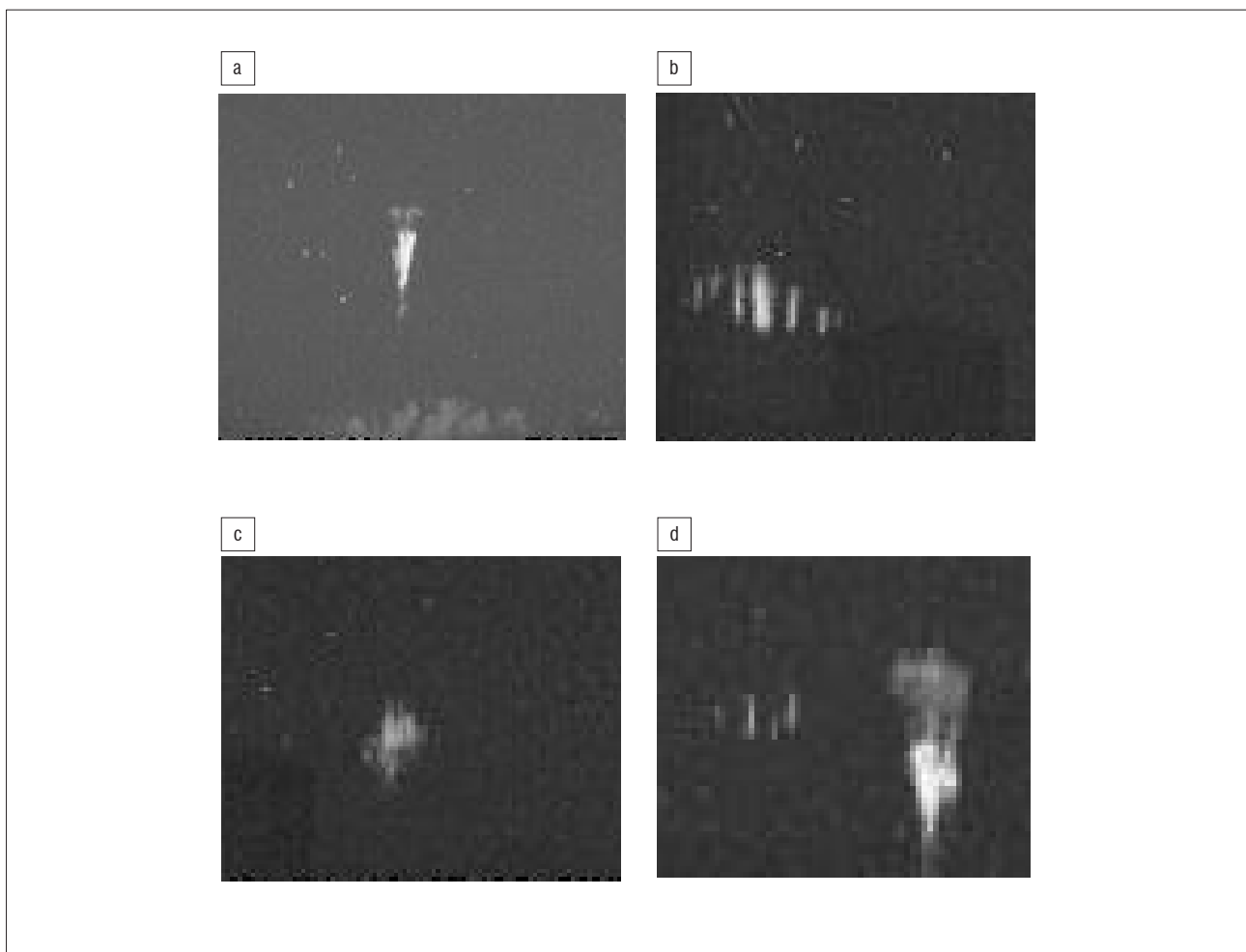


Figure 3: Some examples of the sprite elements recorded on 11 January 2016 with variable zoom: (a) carrot-shaped sprite observed at 18:53:49.4 UTC, (b) seven column-shaped sprites observed at 19:38:32.5 UTC, (c) unclassified sprite observed at 20:08:03.3 UTC, (d) two-carrot/four-column-shaped sprite observed at 19:17:09.8 UTC. Panel (a) shows the very first sprite image that was recorded over southern Africa. The dark square in the bottom left in panels (b)–(d) is a telescope dome in the foreground. The bright spots are stars.

The azimuth uncertainty when locating sprites using simple geometry depends on the distance from the camera to the sprite. For a typical range of 600 km and horizontal pixel resolution of 0.1°, the azimuthal uncertainty translates to ±0.5 km per pixel. The distance (d) is determined from Equation 1 by measuring the elevation angle (θ) to the top of the sprite and assuming an initiation altitude (h) of 85 km.^{9,38,44} The range uncertainty consists of the star-fit image pixel jitter in the vertical direction (1-pixel, which is equivalent to 0.09°) and the sprite initiation altitude (5 km).¹² Again, for a typical range of 600 km, the uncertainty was ±27 km, which is dominated by the uncertainty in the sprite initiation altitude.

Archival lightning stroke data over South Africa were obtained from the South African Weather Service (SAWS). The SAWS operates a network of sensors that detect lightning strokes over South Africa.⁴⁵ The system uses magnetic direction finding and time of arrival of the lightning stroke to determine the actual time and location of the lightning strokes in real time. For post-processing of the sprite events recorded during the observations at 80 ms (2 TV frames), the initiation timestamp on the images together with the coordinates retrieved from the Sonotaco sprite analyser software were compared with the lightning stroke data set from SAWS in position and time, with the assumption that sprites could be initiated within a 50 km distance from the parent lightning stroke^{46,47} and within the 80 ms image temporal resolution.

Distance estimation

Sprites are believed to initiate at altitudes ~85 km^{9,38,44} with maximum brightness around 68 km¹². By measuring the elevation angle, we were able to estimate the distance between the source and observer. The viewing elevation angle was retrieved by star fitting the image frames to many corresponding reference stars in the catalogue. We then derived Equation 2 from Equation 1 in order to solve for the distance (d).

$$d = Re \left[\theta \cdot \sin^{-1} \left(\frac{Re \sin \theta}{h + Re} \right) \right] \quad \text{Equation 2}$$

The results from this calculation show that most of the sprite events occurred approximately between 540 km and 680 km from the observation point. Where the sprites comprised several sprite events in a group, the distance was estimated from the centre of the group. The error in this estimate for one pixel offset in the horizontal and vertical direction was ±5% of the average distance from the observation point.

Sprite brightness estimation

Sprite brightness was estimated by using the ImageJ software (<https://imagej.net/Welcome>). We first averaged the scintillation effects of a star that was present in all the video frames of each captured sprite video clip. This step was done to obtain an average intensity value for the star ($Star_{avg}$) because it has a constant brightness. We then applied a 7 x 7 median filter on all the individual TV frames containing sprites to remove the background stars (Figure 4). Next, we measured the brightness of an area (same height and width) with a sprite element and background sky (Sp+Sky) and also the brightness of two adjacent areas of the sky without sprite elements (Sky_1 , Sky_2 ; marked '1', '2' on Figure 4b) and recorded their average (Sky_{avg}). The essence of using the average was to minimise errors in the estimation of the sky background. Thereafter, we subtracted Sky_{avg} from Sp+Sky to estimate the sky-subtracted brightness of each sprite element (Sp_{bg}). Finally, we normalised the sprite brightness (Sp_{bgl}) by dividing each of these values by the average brightness of the reference star ($Star_{avg}$). Because brightness is inversely proportional to the square of the distance, we finally divided each result by the square of its corresponding distance to the observation point. Supplementary table 1 shows a list of all the sprite events recorded on 11 January 2016, along with their corresponding parameters.

Charge moment change estimation

The charge moment change (CMC) is an important lightning metric for sprite studies that is not measured by the Lightning Detection Networks. It is defined as the product of the charge and the altitude from where these charges were lowered to the ground.⁴⁸ Analogous to a capacitor, the amount of charge and its separation distance determines the voltage and it is the electric field generated by the charge displacement during a lightning strike that triggers sprites. The CMC data were obtained from the Schumann resonance station in Hungary (NCK; 47.62°N, 16.72°E). This station is located ~8947 km away from SAAO. The system records both the vertical electric field components (E_z) and the horizontal magnetic field components (H_{NS} and H_{EW}) associated with lightning strokes.⁴⁹ The CMC associated with some of the sprite events was retrieved from the electric component of the parent lightning strokes as described in Huang et al.⁵⁰ This analysis is simpler when observing the signals in the far field, hence Schumann resonance observations from lightning in South Africa were made in the northern hemisphere. Because of the relatively high noise level in the Schumann resonance system during these observations, only about 20% of the events that were observed were satisfactorily processed.

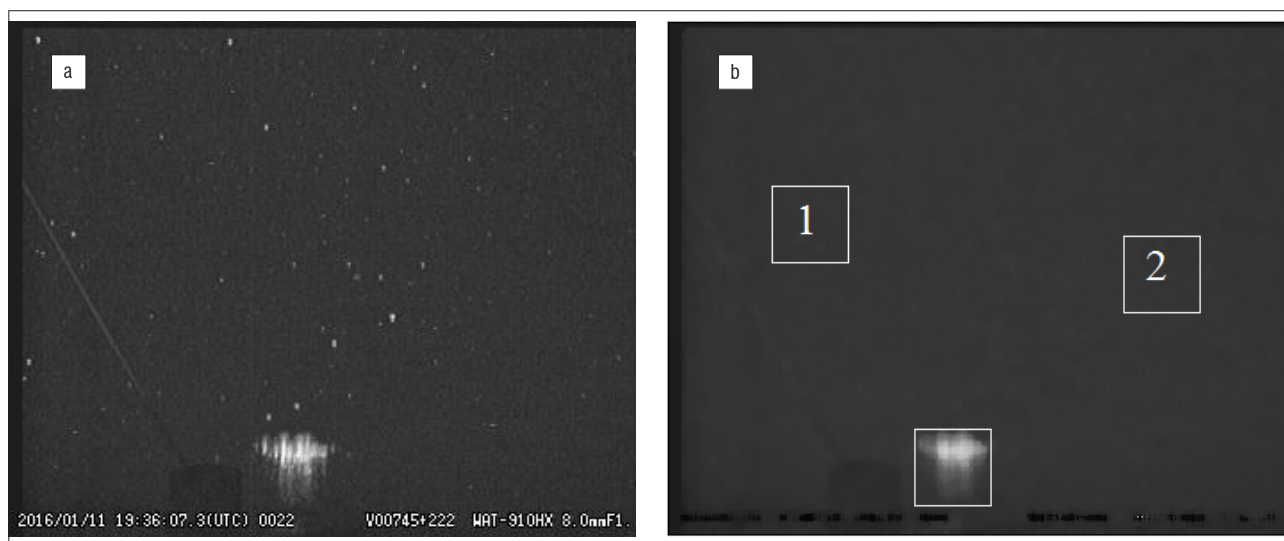


Figure 4: Sprite image (a) before and (b) after applying a 7 x 7 median filter to remove the background stars. The boxes marked '1' and '2' indicate the locations used to derive the sky backgrounds (Sky_1 , Sky_2). The average of these sky backgrounds was used to derive the sky-subtracted brightness of the sprite elements.

The CMC associated with these processed events ranged from 900 C.km to 2000 C.km. These values are consistent with the findings of earlier studies^{51,52} and were in agreement with the theoretical threshold (500–1000 C.km) for dielectric breakdown of the mesosphere³⁵. Figure 5 shows the relationship between sprite brightness, as described above, and charge moment change for all events for which Schumann resonance data and their corresponding parent CG lightning strike were available (Supplementary table 1).

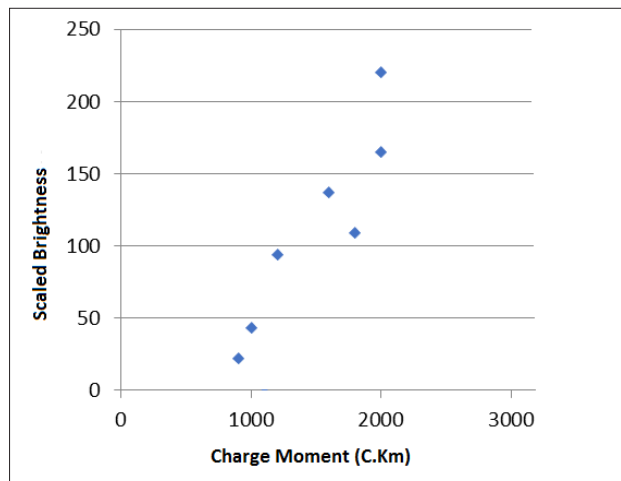


Figure 5: The relationship between the sprite luminosity and the charge moment of the lightning discharges for the events where Schumann resonance data and the parent cloud-to-ground lightning strike were available.

Discussion

As these were the first recorded ground-based images of sprites in southern Africa, we began our analysis with a simple visual inspection of our sprite images to compare them with images recorded in other parts of the world. This comparison showed that the various structures we recorded were approximately the same as those recorded elsewhere. Whilst the physical mechanism(s) that is responsible for the formation of the column and carrot-shaped sprites has been proposed by Qin⁵³, the mechanism that is responsible for other structures of sprites remains poorly understood.

Our analysis also shows that, for the sprite events that were associated with their corresponding lightning strokes, the average lightning peak current associated with these sprite initiations was ~ 74 kA, and that the maximum and minimum current values were 191 kA and 11 kA, respectively. In all cases, the initiating lightning strikes were positive cloud-to-ground lightning strokes (+CGs). These values were derived from the lightning strike location data sets from the SAWS network. The average displacement between each of the sprite elements and the parent lightning flashes was ~ 15 km, which is well within our measurement uncertainty and is in agreement with several earlier studies.^{38,46,47}

A quasi-linear relationship is observed between the CMC and sprite brightness^{37,54} for the events for which their corresponding Schumann resonance data and parent CG lightning strike were available. The larger the CMC, the larger the electric field generated and the greater the optical signature of the resultant gas discharge.

Conclusion

Sprites have been photographically recorded for the first time in southern Africa since the earliest sightings reported anecdotally in 1937. We suggest that sprites in southern Africa have the same morphology as those recorded elsewhere and that the CMC of the recorded events agrees with the threshold for dielectric breakdown of the mesosphere. We have also shown that sprite brightness is related quasi-linearly with CMC, as earlier suggested by Yaniv et al.³⁷ We plan to observe sprites in future using multiple cameras with filters to extract spectral information and estimate the characteristic energy of the electrons within a sprite.

Acknowledgements

We acknowledge the financial assistance of the National Research Foundation (South Africa) through grant no. 105535. We also thank the South African Astronomical Observatory, the South African Weather Service and the Climate System Analysis Group, Department of Environmental and Geographical Science, University of Cape Town, for their support.

Authors' contributions

M.K. was responsible for the conceptualisation, project leadership and project management; S.N., M.K. and P.M. were responsible for the methodology and writing; S.N. and P.M. were responsible for collecting the data; S.N. was responsible for data analysis and sample analysis; J.B. was responsible for charge moment estimation; M.K. and P.M. were responsible for validation, student supervision, funding acquisition.

References

- MacKenzie T, Toyne H. Meteorological phenomena. *Nature*. 1886;33:245. <https://doi.org/10.1038/033245e0>
- Davidson JE. Thunderstorm and aurora phenomena. *Nature*. 1893;47:582. <https://doi.org/10.1038/047582b0>
- Boys CV. Progressive lightning. *Nature*. 1926;118:749–450. <https://doi.org/10.1038/118749a0>
- Wilson CTR. A theory of thundercloud electricity. *Proc R Met Soc Lond*. 1956;236:297–317. <https://doi.org/10.1098/rspa.1956.0137>
- Malan D. Sur les décharges orageuses dans la haute atmosphère [On the stormy discharges in the atmosphere]. Paris: Academie des Sciences; 1937. French.
- Franz RC, Nemezek RJ, Winckler JR. Television image of large upward electrical discharge above a thunderstorm system. *Science*. 1990;249:48–51. <https://doi.org/10.1126/science.249.4964.48>
- Pasko VP, Yair Y, Kuo C-L. Lightning related transient luminous events at high altitude in the earth's atmosphere: Phenomenology, mechanisms, and effects. *Space Sci Rev*. 2012;168:475–516. <https://doi.org/10.1007/s11214-011-9813-9>
- Siingh D, Singh RP, Singh AK, Kumar S, Kulkarni MN, Singh AK. Discharges in the stratosphere and mesosphere. *Space Sci Rev*. 2012;169(1–4):73–121. <https://doi.org/10.1007/s11214-012-9906-0>
- Pasko VP, Inan US, Bell TF, Taranenko YN. Sprites produced by quasi-electrostatic heating and ionization in the lower ionosphere. *J Geophys Res*. 1997;102(A3):4529–4561.
- Stanley M, Krehbiel P, Brook M, Moore C, Rison W. High-speed video of initial sprite development. *Geophys Res Lett*. 1999;26:3201–3204. <https://doi.org/10.1029/1999GL010673>
- Stenbaek-Nielsen HC, Kanmae T, McHarg MG, Haaland R. High-speed observations of sprite streamers. *Surv Geophys*. 2013;34:769–795. <https://doi.org/10.1007/s10712-013-9224-4>
- Sentman DD, Wescott EM, Osborne DL, Hampton DL, Heavner MJ. Preliminary results from the Sprites94 campaign: Red sprites. *J Geophys Res Lett*. 1995;22:1205–1208. <https://doi.org/10.1029/95GL00583>
- Winckler JR, Lyons WA, Nelson T, Nemezek RJ. New high-resolution ground-based studies of cloud-ionosphere discharges over thunderstorms. *J Geophys Res Lett*. 1999;101:6997–7004. <https://doi.org/10.1029/95JD03443>
- Liu NY, McHarg MG, Stenbaek-Nielsen HC. High-altitude electrical discharges associated with thunderstorm and lightning. *J Atmos Sol Terr Phys*. 2015;136:98–118. <https://doi.org/10.1016/j.jastp.2015.05.013>
- Lang TJ, Cummer SA, Rutledge SA, Lyons WA. The meteorology of negative cloud-to-ground lightning strokes with large charge moment changes: Implications for negative sprites. *J Geophys Res Atmos*. 2013;118:7886–7896. <https://doi.org/10.1002/jgrd.50595>
- Boggs LD, Liu N, Splitt M, Lazarus C, Glenn C, Rassoul H, et al. An analysis of five negative sprite-parent discharges and their associated thunderstorm charge structures. *J Geophys Res Atmos*. 2016;121(2):759–784. <https://doi.org/10.1002/2015JD024188>

17. Fukunishi H, Takahashi Y, Uchida A, Sera M, Miyasato KR. Occurrences of sprite and elves above the Sea of Japan near Hukuritku in winter. *EOS Supplement*. 1999;80:217.
18. Neubert T, Rycroft M, Farges T, Blanc E, Chanrion O, Arnone E, et al. Recent results from studies of electric discharges in the mesosphere. *Surv Geophys*. 2008;29:71–137. <https://doi.org/10.1007/s10712-008-9043-1>
19. Sao Sabbas FT, Taylor MJ, Pautet PD, Bailey M, Cummer S, Azambuja RR, et al. Observations of prolific transient luminous event production above a mesoscale convective system in Argentina during the Sprite2006 Campaign in Brazil. *J Geophys Res*. 2010;115, A00E58, 20 pages. <https://doi.org/10.1029/2009JA014857>
20. Hardman SF, Dowden RL, Brundell JB, Bahr JL, Kawasaki Z, Rodger CJ. Sprite observations in the northern territory of Australia. *J Geophys Res*. 2000;105(D4):4689–4697.
21. Blanc E, Farges T, Roche R, Brebion D, Hua T, Labarthe A, et al. Nadir observations of sprites from the International Space Station. *J Geophys Res Atm*. 2004;109, A02306, 8 pages. <https://doi.org/10.1029/2003JA009972>
22. Jehl A, Farges T, Blanc E. Color pictures of sprites from nondedicated observation on board the International Space Station. *J Geophys Res Space Phys*. 2013;118:454–461. <https://doi.org/10.1029/2012JA018144>
23. Sato M, Ushio T, Morimoto T, Kikuchi M, Adachi T, Suzuki M, et al. Overview and early results of the Global Lightning and Sprite Measurements mission. *J Geophys Res Atmos*. 2015;120:3822–3851. <https://doi.org/10.1002/2014JD022428>
24. Cummer SA, Inan US, Bell TF, Barrington-Leigh CP. ELF radiation produced by electrical currents in sprites. *Geophys Res Lett*. 1998;25:1281–1284. <https://doi.org/10.1029/98GL50937>
25. Ignaccolo M, Farges T, Mika A, Allin T, Chanrion O, Blanc E, et al. The planetary rate of sprite events. *J Geophys Res Lett*. 2006;33, L11808, 5 pages. <https://doi.org/10.1029/2005gl025502>
26. Ebert U, Nijdam S, Li C, Luque A, Briels T, Van Veldhuizen E. Review of recent results on streamer discharges and discussion of their relevance for sprites and lightning. *J Geophys Res*. 2010;115, A00E43, 13 pages. <https://doi.org/10.1029/2009JA014867>
27. Pasko VP. Recent advances in theory of transient luminous events. *J Geophys Res*. 2010;115(A6), A00E35, 24 pages. <https://doi.org/10.1029/2009JA014860>
28. Rodger CJ. Red sprites, upward lightning, and VLF perturbations. *Rev Geophys*. 1999;37(3):317–336. <https://doi.org/10.1029/1999RG900006>
29. Hiraki Y, Kasai Y, Fukunishi H. Chemistry of sprite discharges through ion-neutral reactions. *Atmos Chem Phys*. 2008;8:3919–3928. <https://doi.org/10.5194/acp-8-3919-2008>
30. Liu N, Dwyer J, Stenbaek-Nielsen H, McJarg M. Sprite streamer initiation from natural mesospheric structures. *Nat Comm*. 2015;6, Art. #7540, 9 pages. <https://doi.org/10.1038/ncomms8540>
31. Yair Y, Peter I, Adam D, Moalem M, Price C, Joachim H, et al. New observations of sprites from the space shuttle. *J Geophys Res Lett*. 2004;109, D15201, 10 pages. <https://doi.org/10.1029/2003JD004497>
32. Sato M, Fukunishi H. Global sprite occurrence locations and rates derived from triangulation of transient Schumann resonance events. *J Geophys Res Lett*. 2003;30(16), Art. #1859, 4 pages. <https://doi.org/10.1029/2003GL017291>
33. Chen A, Kuo C, Lee Y, Su H, Hsu R, Chern J, et al. Global distributions and occurrence rates of transient luminous events. *J Geophys Res*. 2008;113(A8), A08306, 8 pages. <http://dx.doi.org/10.1029/2008JA013101>
34. Christian H, Blakeslee R, Boccippio D, Boeck W, Buechler D, Driscoll K, et al. Global frequency and distribution of lightning as observed from space by the optical transient detector. *J Geophys Res*. 2003;108(D1):ACL4-1–ACL4-15. <http://dx.doi.org/10.1029/2002JD002347>
35. Williams E, Lyons W, Hobara Y, Mushtak V, Asencion N, Boldi R, et al. Ground-based detection of sprites and their parent lightning flashes over Africa during the 2006 AMMA campaign. *Q J Roy Meteor Soc*. 2010;136(S1):257–271. <https://doi.org/10.1002/qj.489>
36. Soula S, Van der Velde O, Montanya J, Neubert T, Chanrion O, Ganot M. Analysis of thunderstorm and lightning activity associated with sprites observed during the EuroSprite campaigns: Two case studies. *Atmos Res*. 2009;91(2–4):514–528. <https://doi.org/10.1016/j.atmosres.2008.06.017>
37. Yaniv R, Yair Y, Price C, Bor J, Sato M, Hobara Y, et al. Ground-based observations of the relations between lightning charge-moment-change and the physical and optical properties of column sprites. *J Atmospheric Sol-Terr Phys*. 2013;107:60–67. <https://doi.org/10.1016/j.jastp.2013.10.018>
38. Barrington-Leigh CP, Inan US. Identification of sprites and elves with intensified video and broadband array photometry. *J Geophys Res*. 2001;106:1741–1750. <https://doi.org/10.1029/2000JA000073>
39. Lyons WA. Characteristics of luminous structures in the stratosphere above thunderstorm as imaged by a low-light video. *Geophys Res Lett*. 1994;21:875–878. <https://doi.org/10.1029/94GL00560>
40. Kosch M, Pedersen T, Esposito R. Wide angle mirror system design for distortion-less imaging of the sky. *Appl Opt*. 2009;48(24):4703–4708. <https://doi.org/10.1364/AO.48.004703>
41. Collier AB, Hughes AR, Lichtenberger J, Steinbach P. Seasonal and diurnal variation of lightning activity over southern Africa and correlation with European whistler observations. *Ann Geophys*. 2006;24:529–542. <https://doi.org/10.5194/angeo-24-529-2006>
42. Yair Y, Price C, Ganot M, Greenberg E, Yaniv R, Ziv B, et al. Optical observations of transient luminous events associated with winter thunderstorms near the coast of Israel. *Atmos Res*. 2009;91(2–4):529–537. <https://doi.org/10.1016/j.atmosres.2008.06.018>
43. Bor J. Optically perceptible characteristics of sprites observed in Central Europe in 2007–2009. *J Atmospheric Sol-Terr Phys*. 2013;92:151–177. <https://doi.org/10.1016/j.jastp.2012.10.008>
44. Wescott E, Stenbaek-Nielsen H, Sentman D, Heavner M, Moudry D, Sao-Sabbas FT. Triangulation of sprites, associated halos and their possible relation to causative lightning and micrometeors. *J Geophys Res Space Phys*. 2001;106(A6):10467–10477. <https://doi.org/10.1029/2000JA000182>
45. South African Weather Services (SAWS). South African Weather Service Lightning Detection Network (database on the Internet). No date [cited 2018 Jun 06]. Available from: <http://www.weathersa.co.za/observations/lightning>
46. Lyons WA. Sprites observations above the USA high plains in relation to their parent thunderstorm systems. *J Geophys Res*. 1996;101(D23):29641–29652. <https://doi.org/10.1029/96JD01866>
47. Sao Sabbas FT, Sentman DD, Wescott EM, Pinto O, Mendes O, Taylor MJ. Statistical analysis of space-time relationships between sprites and lightning. *J Atmospheric Sol-Terr Phys*. 2003;65(5):525–535. [https://doi.org/10.1016/S1364-6826\(02\)00326-7](https://doi.org/10.1016/S1364-6826(02)00326-7)
48. Cummer SA, Jaugey NC, Li JB, Lyons WA, Nelson TE, Gerken EA. Submillisecond imaging of sprite development and structure. *Geophys Res Lett*. 2006;33(4), L04104, 4 pages. <https://doi.org/10.1029/2005GL024969>
49. Satori G, Szendroi J, Vero J. Monitoring Schumann resonances I: Methodology. *J Atmos Terr Phys*. 1996;58:1475–1481. [https://doi.org/10.1016/0021-9169\(95\)00145-X](https://doi.org/10.1016/0021-9169(95)00145-X)
50. Huang E, Williams E, Boldi R, Lyons W, Taylor M, Nelson T, et al. Criteria for sprites and elves based on Schumann resonance observations. *J Geophys Res*. 1999;104(D14), 16943–16964. <https://doi.org/10.1029/1999JD900139>
51. Hobara Y, Hayakawa M, Williams ER, Boldi R, Downes E. Location and electrical properties of sprite-producing lightning from a single ELF site. In: Fullekrug M, Mareev E, Rycroft M, editors. *Sprites, elves and intense lightning discharges*. NATO science series II: Mathematics, physics and chemistry. Dordrecht: Springer; 2004. p. 211–236. https://doi.org/10.1007/1-4020-4629-4_10
52. Williams ER, Downes E, Boldi R, Lyons WA, Heckman S. Polarity asymmetry of sprite-producing lightning: A paradox? *Radio Sci*. 2007;42, RS2S17, 15 pages. <https://doi.org/10.1029/2006RS003488>
53. Qin J, Celestin S, Pasko VP, Cummer SA, McHarg MG, Stenbaek-Nielsen HC. Mechanism of column and carrot sprites derived from optical and radio observations. *Geophys Res Lett*. 2013;40:4777–4782. <http://dx.doi.org/10.1002/grl.50910>
54. Takahashi Y, Yoshida A, Sato M, Adachi T, Kondo S, Hsu R, et al. Absolute optical energy of sprites and its relationship to charge moment of parent lightning discharge based on measurement by ISUAL/AP. *J Geophys Res*. 2010;115(A9), A00E55, 8 pages. <http://dx.doi.org/10.1029/2009JA014814>

