SOUTH AFRICAN Journal of Science

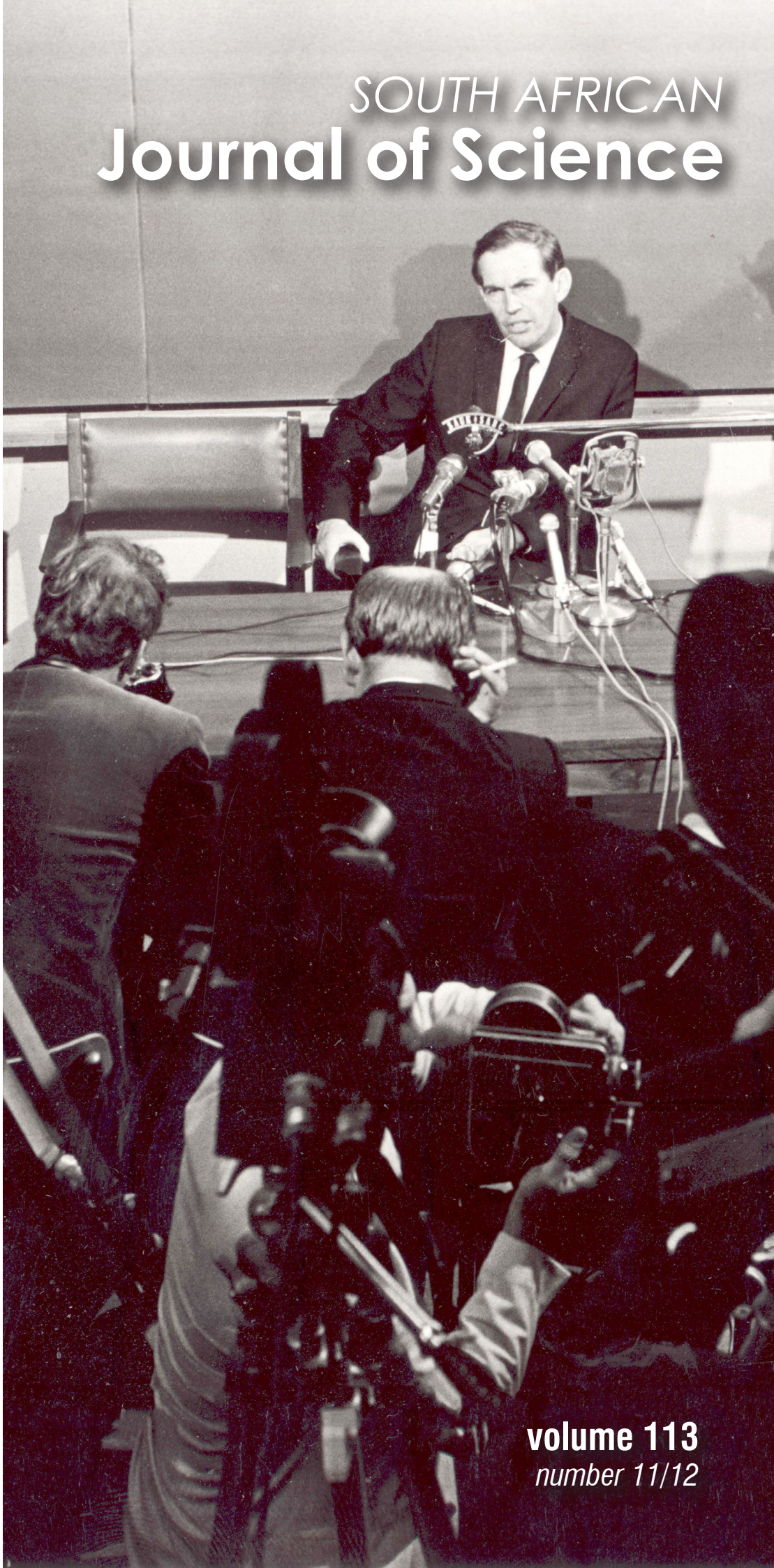
50th anniversary of the first human heart transplant: Reflections on Chris Barnard and celebrity scientists in South Africa today

Environmental impacts of solar power in arid regions of South Africa

South Africa's geothermal energy potential explored

Algae and cyanobacteria for revegetation of gold mine tailings

Seawater quality: Implications for desalination plans for the City of Cape Town



NOVEMBER/DECEMBER 2017

eISSN: 1996-7489

ASSAf
ACADEMY OF SCIENCE OF SOUTH AFRICA

volume 113
number 11/12

NOVEMBER/DECEMBER

eISSN: 1996-7489

SOUTH AFRICAN
Journal of Science

volume 113

number 11/12

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

Submissions
Submissions should be made at http://mc.manuscriptcentral.com/sajs

Cover caption
Chris Barnard at a press conference after performing the first human heart transplant on 3 December 1967 (photo courtesy of the Heart of Cape Town Museum). On this 50th anniversary of that historic event, Joubert examines the life of Chris Barnard and reflects on his subsequent scientific celebrity (p.9) and Joubert and Guenther explore the role of celebrity scientists today and identify the publicly visible scientists in South Africa (p.71).

Research Letter

Avarice: Signs of threats to credible higher education?

Indications of the extent to which higher education is becoming 'commercialised' – driven in a variety of ways by avarice and mostly impelled by external interests (including ranking systems which are not discussed here) – are all around us.

At an individual level, it is evidenced by academics who are committed more to their own interests than those normally associated with scientific practice and the importance of the public good. This avarice takes two forms. Firstly, for individuals, it is the substitution of scientific integrity with personal promotion and subsidy income – evidenced by increases in plagiarism, the publication of research articles in predatory journals (at great expense and without scientific substance or credibility), and the delivery of papers at predatory conferences. Secondly, this individual level of operation serves to fuel the profits of predatory journals and the organisers of predatory conferences.

In 2015, Thomas and De Bruin[1] revealed, in this journal, that plagiarism, driven by a scramble for subsidy payments, was widespread in South African management journals. These authors estimated that the cost to the Department of Higher Education and Training amounted to some ZAR7 million in subsidies paid for unoriginal research. In their paper, Thomas and De Bruin cite a much earlier paper, published in this journal in 2003, titled 'The good, the bad, and the ugly'[2], which includes a definition of scientific 'misconduct':

> *…misconduct by authors in the publication of research results includes misleading authorship, undisclosed conflicts of interest, redundant publication, plagiarism, fabrication of data, selective exclusion of data and breaches of ethical codes.* (p.402)

Almost all such misconduct is driven by self-interest rather than the advancement of science.

Since 2003, further indications of the potential for avarice have emerged, most notably predatory journals and, in the past few years, predatory conferences. Mouton and Valentine[3] have shown just how widespread resorting to predatory journals has become to publish low-quality, often unreviewed or poorly reviewed, research. Their paper reveals that 4245 articles were published in possibly or clearly predatory journals between 2005 and 2014, with a sharp uptick in 2011 leading to a total of 846 such publications in 2014. Notably, the universities with the highest numbers of publications in predatory journals are those with still-developing research histories, including universities of technology.

Predatory conferences

> *do not fulfill the purposes of academic conferences…or mislead attendees about details such as peer review, the for-profit status of the organizers, or the expected number of attendees. …Predatory conferences cut corners by failing to provide proper editorial oversight, for example, by skipping the promised peer review. Some will tout big-name speakers who are not actually involved in the conference.*[4]

Predatory conferences typically also charge substantial attendance fees, with the proceeds going primarily to the organisers. Like many academics, I have been asked frequently to deliver papers at 'high-profile' conferences on subjects about which I know little or nothing!

These forms of profiteering are abetted by varying degrees of misconduct and the pursuit of status and subsidies – representing the 'bad and the ugly' in science and higher education. In so doing, they undermine the credibility of institutions and research – credibility that is very difficult to win back once lost.

At a different level, there are other – more direct – ways in which higher education faces the risks created by profiteering: primarily (and sadly) in the private post-school education sector. The Council for Higher Education, which accredits (or does not accredit) courses, receives applications from a growing number of private higher education institutions in South Africa. These institutions (125 registered or provisionally registered in 2016) range in size from the very small to the very large and serve the educational needs of growing numbers of students. Just under 2 million students were registered in universities and technical and vocational education and training (TVET) colleges at the beginning of last year, of whom 12% were studying in private institutions.

The figure in countries in similar stages of growth as South Africa is often closer to 20% and it is possible that uncertainty in the public higher education sector could see the local figure rise to similar levels. And while some institutions offer sound education (and training), others are borderline profit-making operations – as attested by the observation that, by 2016, 72 private institutions had their registrations cancelled, a further 7 were added to the list in 2016, and 10 requested that their registrations be discontinued – a total of 89 institutions.

But the pressure will also begin to rise in the public institutions: at present, just over a third of institutional income is generated by student fees (the state contribution having decreased from 49% to 39% between 2000 and 2015) and if the fee income is reduced, and state spending remains even at present levels, it will be third-stream income that has to be increased. And that is commercial income in various forms.

Amongst the most egregious examples of private higher education institutions that fail to deliver credible education (or any education at all) are found in the USA. Donald Trump's private university in San Diego was ordered to pay USD25 million to aggrieved students, and there was a threat that Trump might be tried on criminal racketeering charges. Betsy DeVos, Trump's Education Secretary (who understands little about education) has hired as a special advisor Robert Eitel, formerly Chief Compliance Officer (and Vice President) for Bridgeport Education Inc., a for-profit private institution which has faced (and faces) multiple government investigations, the closure of one of its campuses and a settlement of USD30 million because of deceptive student lending. In a decision equally inspired by high levels of corporate integrity, DeVos has appointed Julian Schmoke to head her office that oversees fraud. Schmoke was, until his appointment to government, a Dean at DeVry University which recently settled a USD100 million lawsuit for misleading students. A case of appointing commercial wolves to take care of the public sheep.

There are aspects of the commercialisation of discoveries that are undoubtedly essential to the survival of all universities – but those that are directed at personal gain, profiteering or fraud (directed at students or those who gradually come to direct practices in research and teaching) are certainly potential threats to the broader credibility of higher education.

References

1. Thomas A, De Bruin GP. Plagiarism in South African management journals. S Afr J Sci. 2015;111(1/2), Art. #2014-0017, 3 pages. http://dx.doi.org/10.17159/sajs.2015/20140017
2. The good, the bad, and the ugly. S Afr J Sci. 2003;(9/10):402–403.
3. Mouton J, Valentine A. The extent of South African authored articles in predatory journals. S Afr J Sci. 2017;113(7/8), Art. #2017-0010, 9 pages. http://dx.doi.org/10.17159/sajs.2017/20170010
4. Zepernick J, Musick C. How to avoid predatory conferences. Think Science. 2017 July 26. Available from: https://thinkscience.co.jp/en/articles/avoiding-predatory-conferences.html

HOW TO CITE: Butler-Adam J. Avarice: Signs of threats to credible higher education? S Afr J Sci. 2017;113(11/12), Art. #a0246, 1 page. http://dx.doi.org/10.17159/sajs.2017/a0246

A call to halt destructive, illegal mining in Zimbabwe

AUTHORS:
Mwazvita T.B. Dalu[1]
Ryan J. Wasserman[2]
Tatenda Dalu[3]

AFFILIATIONS:
[1]Environmental Science Department, Rhodes University, Grahamstown, South Africa
[2]School of Science, Monash University Malaysia, Selangor Darul Ehsan, Malaysia
[3]Department of Zoology and Entomology, Rhodes University, Grahamstown, South Africa

CORRESPONDENCE TO:
Tatenda Dalu

EMAIL:
dalutatenda@yahoo.co.uk

KEYWORDS:
artisanal mining; aquatic ecosystem vulnerability; biodiversity loss; ecosystem service loss; river degradation; sedimentation

HOW TO CITE:
Dalu MTB, Wasserman RJ, Dalu T. A call to halt destructive, illegal mining in Zimbabwe. S Afr J Sci. 2017;113(11/12), Art. #a0242, 2 pages. http://dx.doi.org/10.17159/sajs.2017/a0242

Amidst the socio-economic meltdown in Zimbabwe, increased illegal mining activities on recently discovered diamond and gold deposits have given local people a source of livelihood. The problem of illegal gold mining is fast spreading through the Eastern Highlands and other parts of the country, as hundreds of artisanal gold miners swarm to the area, tearing up natural forests and plantations. Illegal mining operations are now rampant in the Chimanimani Mountains, a narrow belt located on the eastern Zimbabwean/western Mozambican border. The hilly and generally inaccessible nature of these mountains has preserved much of the native flora and fauna, including some endemic or near endemic animals and plants,[1] and thus they are an important habitat. The highlands are also a crucial international catchment and a major water source for thousands of families dependent on agriculture and fishing for their livelihoods. However, with an estimated 400 000 miners now operating illegally in Zimbabwe,[2] the cumulative mining activities represent a major threat to people's livelihood security and to the integrity of this crucial catchment, habitat and associated river systems.

Classical artisanal mining in the region (banned in Zimbabwe since 2006) involved the excavation of river bottom sediment for the extraction of gold. River sediment was then processed for the separation of the precious heavy metal from non-target material. More recently, however, many small-scale artisanal activities have evolved into medium- and even large-scale operations, with new processes and/or techniques that involve diverting entire streams/rivers into larger areas holding alluvial gold deposit traces. This water is used to erode the off-stream sediment before being re-diverted into constructed streams which run through makeshift sorting stalls which trap gold. While the illegal artisanal mining engineering ingenuity is a marvel, its trade-off with the environment is unreasonably high. In many instances, harmful chemicals such as mercury and cyanide are also employed in the extraction process.[3,4] We recorded mercury concentrations of 0.1–0.3 mg/kg in river sediment within the illegal mining areas, compared with 0 mg/kg mercury where there was no mining. However, there are far broader occupational safety and public health issues that deserve attention, and have resulted in this practice being considered high risk because of the many individuals who undertake it out of livelihood necessity. Smith et al.[5] highlighted the challenges to mitigating health and safety risks among regulatory agencies that work with limited resources to meet the inspection requirements and concluded that research on health and safety issues in the artisanal mining sector must incorporate a regional and holistic approach with both quantitative and qualitative data collection methodologies. We observed that poor ventilation in tunnels and other risky mining practices by these illegal gold miners poses significant health and safety impacts to the miners. We speculate that the required air quantity in underground mining operations will be very low considering that the number of miners can be high, thus increasing the demand for airflow. Most of these miners do not wear protective clothing or equipment (e.g. dust masks), which poses significant risk for dust and chemical exposure for the miners. Further, the illegal miners work prolonged hours (sometimes >24 h at a time), which means they are exposed to chemicals and dust for long periods. Nearby communities can also be exposed to dust and chemicals.

In addition to the damage that these artisanal mining practices impose upon personal health, safety and well-being, these practices also result in on-site river, riparian and terrestrial ecosystems' destruction through sediment and chemical pollution, along with severe and negative alterations to the rivers' banks and courses. Extensive riparian and other terrestrial vegetation removal has resulted in considerable soil loss through erosion. Near endemics such as the freshwater crab *Potamonuates mutareensis* and other aquatic organisms are seriously affected with no life being observed in affected sections of the river systems.[1] Illegal gold diggers are washing away tonnes of soil into rivers at great loss to aquatic biodiversity. These activities are not only destroying vast portions of both terrestrial and aquatic ecosystems, but are also threatening the livelihood of thousands of villagers along the river valleys where these operations are concentrated. For example, the herbs that used to be collected from the river banks by rural communities, to prepare natural home-made remedies for mild ailments, are fast disappearing and/or no longer growing along the rivers because of heavy pollution by dangerous chemicals that the gold miners use, such as mercury and cyanide.

It is for these reasons that the Zimbabwean government urgently needs to revisit the monitoring and restrictions of artisanal mining. While the magnitude of the problem has not gone unnoticed by officials, the criminalisation of artisanal mining in 2006[6] has largely failed to stop this poverty-driven activity. It is our view that the most pervasive and systematic threats to Zimbabwe's and the region's biodiversity as a whole are rooted in poor monitoring, management and legislation as institutional frameworks governing mining are normally sensitive only to formal mining operations and are blind to rapidly expanding illegal artisanal mining.[7] An evaluation of the associated risks, an appraisal of mining laws and monitoring protocols as well as better enforcement of current laws are required for the adequate protection of the environment and local communities. Alternative forms of control should also be considered, such as the formalisation of more sustainable small-scale practices that conform to artisanal rather than fully commercial ventures. Formalisation of artisanal mining will also assist in achieving good governance, transparency of mineral production and trade while ensuring that social requirements are met and the environment is protected. We further suggest tightening the controls on mercury imports so that the country can meet its 2013 *Minamata Convention on Mercury* signatory commitment. Although the legalisation of artisanal mining can be considered a means to livelihood security for some, for many still depending on the natural terrestrial and aquatic systems, its monitoring and restrictions need to be better managed for all stakeholders to ensure a win–win scenario for the miners, environment, local communities and government.

Acknowledgements

This study was funded by the Rufford Foundation (grant no. 21054-2 to T.D.). Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors, and the Rufford Foundation does not accept any liability in this regard.

References

1. Dalu T, Sachikonye MT, Alexander ME, Dube T, Froneman WP, Manungo KI, et al. Ecological assessment of two species of potamonautid freshwater crabs from the eastern highlands of Zimbabwe, with implications for their conservation. PLoS ONE. 2016;11, e0145923, 17 pages. https://doi.org/10.1371/journal.pone.0145923
2. Metcalf SM, Veiga MM. Using street theatre to increase awareness of and reduce mercury pollution in the artisanal gold mining sector: A case from Zimbabwe. J Clean Prod. 2012;37:179–184. https://doi.org/10.1016/j.jclepro.2012.07.004
3. Hall SL. Talk of "clean coal" ignores dirty truth. Front Ecol Environ. 2009;7(2):70. https://doi.org/10.1890/09.WB.001
4. Saldarriaga-Isaza A, Villegas-Palacio C, Arango S. The public good dilemma of a non-renewable common resource: A look at the facts of artisanal gold mining. Resour Policy. 2013;38(2):224–232. https://doi.org/10.1016/j.resourpol.2013.02.001
5. Smith NM, Ali S, Bofinger C, Collins N. Human health and safety in artisanal and small-scale mining: An integrated approach to risk mitigation. J Clean Prod. 2016;129:43–52. https://doi.org/10.1016/j.jclepro.2016.04.124
6. Spiegel SJ. Shifting formalization policies and recentralizing power: The case of Zimbabwe's artisanal gold mining sector. Soc Nat Resour. 2015;28(5):543–558. https://doi.org/10.1080/08941920.2015.1014606
7. Dube N, Moyo F, Sithole M, Ncube G, Nkala P, Tshuma N. Institutional exclusion and the tragedy of the commons: Artisanal mining in Matabeleland South Province, Zimbabwe. Extract Industr Soc. 2016;3(4):1084–1094. https://doi.org/10.1016/j.exis.2016.08.006

A cladistic analysis of *Graecopithecus*

AUTHORS:
Julien Benoit[1]
Francis J. Thackeray[1]

AFFILIATION:
[1]Evolutionary Studies Institute, School of Geosciences, University of the Witwatersrand, Johannesburg, South Africa

CORRESPONDENCE TO:
Julien Benoit

EMAIL:
julien.benoit@wits.ac.za

KEYWORDS:
Hominini; Africa; palaeobiogeography

HOW TO CITE:
Benoit J, Thackeray FJ. A cladistic analysis of *Graecopithecus*. S Afr J Sci. 2017;113(11/12), Art. #a0238, 2 pages. http://dx.doi.org/10.17159/sajs.2017/a0238

Fuss et al.[1] have recently claimed that the earliest Hominini – and incidentally the whole evolutionary root of humankind – could be found, not in Africa, but in Europe. This claim was critically discussed by the media in a number of articles[2,3].

This new hypothesis was proposed on the basis of the re-assessment of a mandible and lower fourth premolars (p_4) from Greece and Bulgaria attributed to *Graecopithecus freybergi* and cf. *Graecopithecus* sp., respectively.[1] Based on an X-ray tomographic revision of this material, Fuss et al.[1] found that *Graecopithecus* shares with Hominini a partial fusion of the p_4 buccal roots (a state that resembles Tomes' root in modern humans), thick enamel and megadonty. *Graecopithecus* fossils were dated to around 7.2 million years, which would make them the oldest remains attributed to Hominini, even older than *Sahelanthropus tchadensis* dated at around 6–7 million years.

However, none of the characters cited by Fuss et al.[1] is strictly unique to Hominini, as thick enamel and megadonty are found in a wide variety of Miocene apes as well as in extant *Pongo*.[4,5] A partial fusion of p_4 roots is present in 2–5% of *Pan* specimens, as acknowledged by Fuss et al.[1] These findings put the assertion that *Graecopithecus* belongs to Hominini into serious question.

Even if *Graecopithecus* can be attributed to Hominini, the fact that it is older than *Sahelanthropus* does not make it the basal-most representative of this clade. As recently exemplified by the *Homo naledi* case[7], the stratigraphic age of a fossil taxon is not a reliable indicator of its phylogenetic position[8]. Fuss et al.[1] emphasised the fact that *Graecopithecus* appears to be more derived than *Sahelanthropus,* both in terms of canine reduction and the degree of p_4 roots fusion. If *Graecopithecus* happens to be more derived than *Sahelanthropus*, then the evolutionary tree of Hominini would remain rooted in Africa and *Graecopithecus* would only represent an offshoot that dispersed out of Africa very early in the evolutionary history of hominins. On the other hand, *Graecopithecus* might be closely related to *Ouranopithecus*, with which it has been synonymised for a long time[9] or to other Eurasian apes, as suggested by previous cladistic analyses[10]. In these cases, the evolutionary root of humankind would definitely remain in Africa.

The re-attribution of *Graecopithecus* by Fuss et al.[1] constitutes an important taxonomic and phylogenetic assertion that has critical implications regarding the early evolutionary origin of Hominini. This assertion must be tested using a cladistic analysis as it provides a standardised method that enables one to reconstruct character polarity and tree topology in a repeatable and testable manner.[11,12] The aim of this short paper is to assess the phylogenetic position of *Graecopithecus* using a cladistic analysis and to discuss the biogeography of early hominins.

We used Finarelli and Clyde's character matrix[4], which is itself an updated version of that of Begun et al.[13], which is the most comprehensive character matrix available that includes Miocene apes and Hominini. *Graecopithecus* was coded following the description by Fuss et al.[1] Two characters were added to the matrix in order to account for the discovery of new diagnostic features in *Graecopithecus*: first, the thickness of the enamel and second, the fusion of the p_4 roots (see the supplementary material). These two characters were coded using previous reports in the literature.[1,6,14] As stated by Fuss et al.[1], fusion of the p_4 buccal roots sometimes occurs in *Pan*; however, in order to reflect the rarity of this condition, this character was coded as absent in this taxon. It must be noted that to code this character as variable in *Pan* does not change the results of the cladistic analysis.

The analysis was run using *Proconsul* as the outgroup. The data matrix was treated under the assumption of the minimal model of unweighted parsimony, using PAUP.4b1,[15] with a branch-and-bound search (an exhaustive search). All characters were treated as unordered and equally weighted. The data matrix is provided in the supplementary material.

The analysis resulted in 15 equally parsimonious trees of 439 steps. The homoplasy index is 0.45, the retention index is 0.64 and the consistency index (CI) is 0.55. The strict consensus is unresolved for the clade unifying Hominini, *Graecopithecus*, *Pan*, *Gorilla*, *Pongo*, *Sivapithecus*, *Lufengpithecus*, *Ouranopithecus* and *Ankarapithecus*. Therefore, only the majority consensus is presented in Figure 1. This tree supports a close relationship between Hominini, *Pan* and *Gorilla*, to the exclusion of *Graecopithecus*, therefore rooting the evolutionary origin of humankind in Africa. *Graecopithecus* appears to be in an unresolved position. Nevertheless, among the 15 equally parsimonious trees, *Graecopithecus* appears as the sister taxon of Hominini in four of them. Two characters support this relationship, but they are both subject to homoplasy:

- 100 (changes to state 2, ambiguous, CI: 0.50): reduced canine
- 113 (changes to state 1, ambiguous, CI: 0.33): postcanine dentition larger than anterior dentition

In one of the four trees, there are two additional synapomorphies:

- 112 (changes to state 1, ambiguous, CI: 0.25): low dentine penetrance
- 201 (changes to state 1, ambiguous, CI: 0.33): thickened enamel

The presence of Tome's root (character 202) appears as an unambiguous synapomorphy only in accelerated transformation optimisation (ACCTRAN scenario). In this respect, it is noticeable that numerous authors have emphasised the importance of homoplasies in Miocene apes, including dental morphology.[6,16] This analysis highlights some of the characters recognised by Fuss et al.[1] to identify *Graecopithecus* as a Hominini in some of the phylogenetic trees, but remarkably reconstructs none of them as a definite, unambiguous synapomorphy.

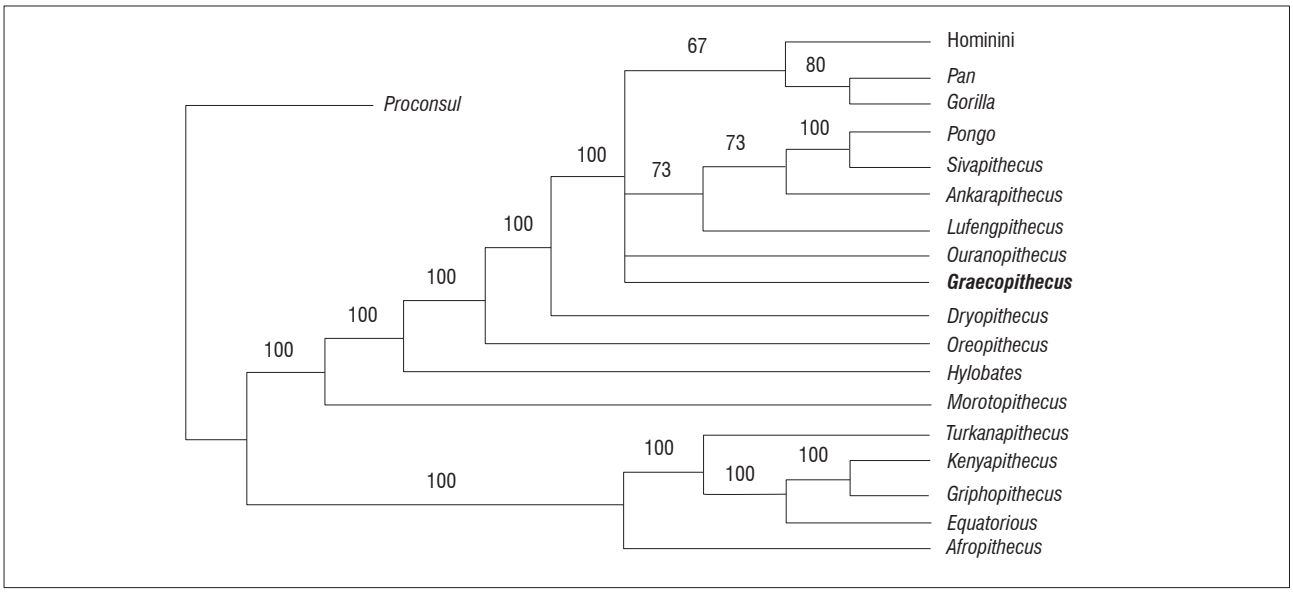


Figure 1: Majority rule consensus of the cladistic analysis performed in this study. Numbers on branches indicate the frequency (%) of the clade among the 15 most parsimonious trees. *Graecopithecus* appears in an unresolved position which indicates that it is either reconstructed as a Ponginae or a Homininae.

With 4 trees among 15 supporting a sister-group relationship between Hominini and *Graecopithecus*, we recognise a small signal for placing *Graecopithecus* at the root of the Hominini clade. This means that the phylogenetic relationship between *Graecopithecus* and Hominini is as yet not confirmed. Our analysis supports the view that *Graecopithecus* is potentially an important taxon for the origin of Hominini, but this is not certain and deserves further investigation and more material.

Acknowledgements

We thank the National Research Foundation (South Africa), the Claude Leon Foundation, the DST-NRF Centre of Excellence for the Palaeosciences, and the Andrew W. Mellon Foundation for support.

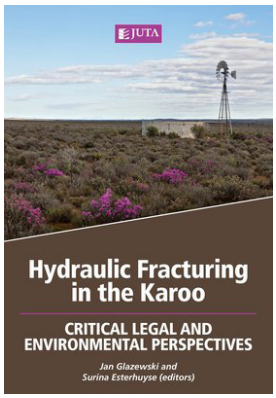
References

1. Fuss J, Spassov N, Begun DR, Böhme M. Potential hominin affinities of *Graecopithecus* from the Late Miocene of Europe. PLoS ONE. 2017;12(5), e0177127, 23 pages. https://doi.org/10.1371/journal.pone.0177127
2. Benoit J. There's not enough evidence to back the claim that humans originated in Europe. The Conversation Africa. 2017 May 25; Science and Technology. Available from: https://theconversation.com/theres-not-enough-evidence-to-back-the-claim-that-humans-originated-in-europe-78280
3. Hawks J. Features of the Grecian ape raise questions about early hominins. John Hawks Weblog. 2017 May 22. Available from: http://johnhawks.net/weblog/fossils/miocene/graecopithecus/graecopithecus-fuss-2017.html
4. Finarelli JA, Clyde WC. Reassessing hominoid phylogeny: Evaluating congruence in the morphological and temporal data. Paleobiology. 2004;30:614–651. https://doi.org/10.1666/0094-8373(2004)030<0614:RHPECI>2.0.CO;2
5. Moyà-Solà S, Köhler M, Alba DM, Casanovas-Vilar I, Galindo J. *Pierolapithecus catalaunicus*, a new Middle Miocene great ape from Spain. Science. 2004;306(5700):1339–1344. https://doi.org/10.1126/science.1103094
6. Begun DR. Fossil record of Miocene hominoids. In: Henke W, Tattersall I, editors. Handbook of paleoanthropology. Berlin: Springer; 2015. p. 1261–1332.
7. Berger LR, Hawks J, Dirks P, Elliott M, Roberts EM. *Homo naledi* and Pleistocene hominin evolution in subequatorial Africa. eLife. 2017;6:e24234, 19 pages. https://doi.org/10.7554/eLife.24234
8. Hawks J, Berger LR. The impact of a date for understanding the importance of *Homo naledi*. Trans Roy Soc S Afr. 2016;71:125–128. http://dx.doi.org/10.1080/0035919X.2016.1178186
9. Begun DR. European hominoids. In: Hartwig WC, editor. The primate fossil record. Cambridge, UK: Cambridge University Press; 2002. p. 339–407.
10. Begun DR. Dryopithecins, Darwin, de Bonis, and the European origin of the African apes and human clade. Geodiversitas. 2009;31:789–816. https://doi.org/10.5252/g2009n4a789
11. Darlu P, Tassy, P. La reconstruction phylogénétique. Concepts et methods [Phylogenetic reconstruction. Concepts and methods]. Paris: Masson et Cie; 1993. French.
12. Wiley EO, Siegel-Causey D, Brooks DR, Funk VA. The compleat cladist: A primer of phylogenetic procedures (special publication, no. 19). Lawrence, KS: University of Kansas Museum of Natural History; 1991.
13. Begun DR, Ward CV, Rose MD. Events in hominoid evolution. In: Begun DR, Ward CV, Rose MD, editors. Function, phylogeny, and fossils: Miocene hominoid evolution and adaptations. New York: Plenum Press; 1997. p. 389–415.
14. Alba DM, Fortuny J, Moyà-Solà S. Enamel thickness in the Middle Miocene great apes *Anoiapithecus*, *Pierolapithecus* and *Dryopithecus*. Proc R Soc B. 2010;277(1691):2237–2245. https://doi.org/10.1098/rspb.2010.0218
15. Swofford DL. PAUP*: Phylogenetic analysis using parsimony (and other methods). Version 4. Sunderland, MA: Sinauer Associates; 2002.
16. Delezene LK, Lucas K, Schrein CM. Is *Ouranopithecus macedoniensis* a hominin? A critical assessment of evidence from canine and P_3 morphology. In: Abstracts of the Annual Meeting of the American Association of Physical Anthropologists; 2005 April 6–9; Milwaukee, WI, USA. A10.

Critical questions (and some answers) in debates on shale gas

BOOK TITLE:
Hydraulic fracturing in the Karoo: Critical legal and environmental perspectives

BOOK COVER:



EDITORS:
Jan Glazewski and Surina Esterhuyse

ISBN:
9781485118183 (softcover)

PUBLISHER:
Juta and Company, Cape Town; ZAR650

PUBLISHED:
2016

REVIEWERS:
Paul Hardcastle[1]
Gerhard Gerber[1]
Henri Fortuin[1]

AFFILIATION:
[1]Western Cape Department of Environmental Affairs and Development Planning, Cape Town, South Africa

CORRESPONDING REVIEWER:
Paul Hardcastle

EMAIL:
Paul.Hardcastle@westerncape.gov.za

HOW TO CITE:
Hardcastle P, Gerber G, Fortuin H. Critical questions (and some answers) in debates on shale gas. S Afr J Sci. 2017;113(11/12), Art. #a0245, 1 page. http://dx.doi.org/10.17159/sajs.2017/a0245

The level of dialogue and readiness in terms of regulatory and governance systems in South Africa has come a long way since 2011, when applications for shale gas exploration were first submitted for large parts of the Karoo Basin. Various statutory and administrative measures have been instituted since then; in addition, a Strategic Environmental Assessment for shale gas development in the Karoo and a Technical Readiness Study were completed to support policy reform. While more information has become available over the years, many answers are still to be provided. This book assists by posing critical questions and factors that must be taken into account.

Part 1 of the book deals with the relevant legislative and governance context and provides a summary of the key statutes that will regulate shale gas development. It further highlights cooperative governance across the three spheres of government as a Constitutional principle and one of the biggest governance challenges. The authors advocate a precautionary approach to safeguard the environmental right of the public. Interestingly, one of the views expressed in the book was vindicated by a High Court decision (*Stern NO and Others v Minister of Mineral Resources* (5762/2015) [2017] ZAECGHC 109) issued on 17 October 2017, which set aside the fracking regulations because the environmental content should have been promulgated in terms of environmental legislation by the Minister of Environmental Affairs.

In Part 2, South Africa's energy context is discussed, to establish whether there is a case for fracking. It examines South Africa's historical and current energy mix, its plans for meeting future energy demand and the potential contribution from shale gas, and also assesses shale gas in relation to other energy sources. The conclusion is that a proper evaluation of shale gas in South Africa's energy mix will require more detailed information, as well as a coherent, rational and sustainable national energy policy, which is lacking at present.

Regarding the economics of shale gas, the potential cost and impact of shale gas production is unknown. The authors argue that the South African experience is more likely to mirror that of countries like China and the United Kingdom, and they draw on the North American experience where much of the international evidence lies. They caution that an independent investigation of the claimed economic benefits and environmental externalities of fracking is required for government to make informed decisions on fracking and the future of shale gas.

The chapter on revenue and tax policy considerations looks at striking an appropriate balance between risk and reward for both government and investor(s). It concludes that shale gas projects have peculiarities that are best harnessed with fiscal regimes that have been employed in other jurisdictions like Poland and Norway, and that corporate income tax rates and/or royalty tax for the shale gas industry should be looked at afresh.

Part 3 of the book deals with geohydrology, water, biodiversity, and archaeology and palaeontology. Waste management and air quality, in particular, would have benefitted from a more in-depth discussion, beyond just a mere mention thereof in the discussion on regulatory context. The various chapters provide sound scientific information to contextualise the debates and concerns for the reader. Most prominent amongst these is the concern about the life-supporting surface and subsurface water resources of the Karoo. The lack of knowledge about the subsurface environment is highlighted and the authors call for the application of the precautionary principle. The book highlights that globally, and even more so in South Africa, there is a lack of information and understanding of the ecological implications of shale gas development, hence the need for baseline assessments and continued monitoring. The chapter on archaeology and palaeontology illustrates why these resources are internationally recognised and of socio-economic value. It emphasises the need for increased research and appropriate legal protection measures.

Part 4 looks at public health, sociology and humanities. The chapter on public health highlights why a precautionary approach is advocated, based on experiences in other parts of the world. The chapter dealing with social fabric, local livelihoods and social psyche highlights particular socio-economic vulnerabilities in the communities of the Karoo, and that great care must be taken to ensure that those most in need do not reap a disproportionate share of the disadvantages brought about by shale gas development. The chapter on evidence-based research in decision-making, as well as the one on the rhetoric of fracking in South Africa, brings to the fore the fundamental importance of appropriate framing and how important it is to ask the right questions. The insights gained from this chapter are so fundamental to the framing of issues on shale gas development and the content of the book, that it is strongly recommended that this chapter be read after the introduction. The chapters on sense of place and planning perspectives – which deal with values, ethics and responsibility – are closely linked and serve as a reminder that decisions on shale gas development must not only be procedurally and legally compliant, but also justifiable in light of broader imperatives, values and ethical considerations.

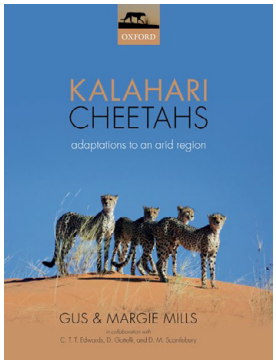
Part 5 is the conclusion to the book; in this regard, we believe that ultimately a decision on shale gas will have to meet the substantive test of having to secure *ecological sustainable* development while promoting *justifiable* economic and social development. The book highlights the complexity of the issues involved in getting to such a decision. We are of the opinion that it is important to differentiate between potential negative consequences requiring more certainty, and those that present fatal flaws, as well as between the need for avoidance on the one hand and better management on the other. This should be facilitated by a phased approach towards decision-making, that is, first allow only conventional exploration then, based on the findings thereof, allow for unconventional exploration (i.e. fracking) and then, based on the findings from the latter, allow for shale gas production, where appropriate.

This book should be read by everyone involved with, or interested in, shale gas development. Its value lies in pulling together the best available legal and environmental information, to inform the critical questions that must be considered in debates on shale gas development.

A new look at cheetahs

BOOK TITLE:
Kalahari cheetahs: Adaptations to an arid region

BOOK COVER:



AUTHORS:
Gus Mills and Margie Mills

ISBN:
9780198712145 (hardcover)

PUBLISHER:
Oxford University Press, Oxford; GBP65

PUBLISHED:
2017

REVIEWER:
Brian W. van Wilgen

AFFILIATION:
Centre for Invasion Biology, Department of Botany and Zoology, Stellenbosch University, South Africa

EMAIL:
bvanwilgen@sun.ac.za

HOW TO CITE:
Van Wilgen BW. A new look at cheetahs. S Afr J Sci. 2017;113(11/12), Art. #a0236, 1 page. http://dx.doi.org/10.17159/sajs.2017/a0236

One of the many advantages of large protected areas is that they offer the opportunity for systematic study of sparsely distributed, wide-ranging species that are fast disappearing from most other areas. The cheetah, which once occurred widely across Africa, the Middle East and Asia, is one such species. Just about all that is known of the ecology of this iconic species comes from studies in two areas – Tanzania's Serengeti National Park and ranching areas in Namibia. The three million hectare Kgalagadi Transfrontier Park, shared between South Africa and Botswana, supports a healthy population of cheetahs, which coexist alongside a host of other large predators (lions, leopards and hyaenas) in a relatively arid region, offering the opportunity to gain new insights into the ecology, behaviour and survival of cheetahs.

There is arguably no one better qualified to conduct such a study than Gus and Margie Mills. In 1972, as newlyweds, the couple arrived in the Kalahari to spend a couple of years studying the little-known brown hyaena. They ended up spending 12 years in the area – a tale entertainingly recounted in a popular book of their adventures.[1] During this time, Gus also produced an authoritative and unique account of the comparative behavioural ecology of both brown and spotted hyaenas.[2] After completing his work in the Kalahari, Gus moved on to the Kruger National Park where he studied the ecology and conservation of wild dogs and lions. During a career as a research scientist that lasted until 2006, he established himself as a world leader in the field of carnivore ecology and conservation. Following 'retirement' from the South African National Parks, the intrepid couple returned to the Kalahari where – with funding from the Lewis Foundation and other sources – they were able to study the ecology of cheetahs for 6 years.

The authors used a combination of good, old-fashioned field observation (over 7000 hours in total, assisted by San trackers) and modern technology to reveal the fascinating lives of these top carnivores. For example, daily energy expenditure was measured using doubly labelled water; tri-axial accelerometers, GPS loggers and drop-off radio collars were used to apportion different energy costs to different behaviours; and DNA studies were used to track paternity and genetic relationships between individuals. The emergence of affordable digital photography allowed the authors to harness the collective observational capacity of hundreds of tourists. By inviting them to submit photographs of cheetahs, they obtained over 1200 batches of photographs, allowing Margie to identify 216 individual cheetahs over 7 years. All of this work has confirmed the status of the cheetah as a sprinting specialist, but it has also revealed important differences between the Kalahari and the better-known Serengeti cheetahs, and provided new insights that have wider implications for conservation.

This book reveals, once again, that the widely held view that the lion is the only social cat is incorrect. Male cheetahs often range alone, but are just as likely to team up in duos or trios. These male coalitions are not necessarily only between siblings, and the coalition males weighed more than single males as they are able to bring down larger prey, and thus eat more. Coalitions, however, do not necessarily confer an evolutionary advantage. Female cheetahs mate with several male individuals, both singleton and coalition males, and litters of cubs typically have more than one father. In the Serengeti, only 5% of cubs survived to adolescence, and many were killed by lions or hyaenas, whereas a third of cubs survived in the Kalahari. Although some were killed by larger predators, it seems possible that many were killed by smaller predators, such as ratels or jackals. The Kalahari cheetahs' main prey was the steenbok, followed by the springbok, with gemsbok, ostrich and even eland calves, as well as hares and the strictly nocturnal springhare forming part of their diet. Cheetahs in the Kalahari live for about 7 years, after which they are either killed (sometimes by other cheetahs), or succumb to starvation.

The book concludes by examining the pressing question of cheetah conservation. Prior understanding was based on the view that cheetahs require vast areas to maintain viable populations, both because they were thought to be limited by competition from other large predators, and because of a lack of genetic variability which placed smaller populations at undue risk. Consequently, international cheetah conservation efforts were focused on maintaining populations outside of protected areas. The Kalahari study has revealed that this view does not necessarily hold. Cheetahs were found not to be at undue risk from other predators, and their genetic makeup has not placed them at any disadvantage – in fact, cheetahs that were re-introduced to protected areas much smaller than the Kgalagadi have thrived. As conservation funds are limited, the authors argue that conservation efforts should focus on maintaining or improving the integrity of protected areas rather than trying to conserve cheetahs in unprotected areas.

I found this book extremely interesting, loaded as it is with facts and illustrations about the diet, hunting behaviour, breeding and survival of cheetahs. The authors have made a valuable contribution to our understanding of the behaviour and survival of a top predator, and in so doing have joined an elite band of eminent authors in this field, including George Schaller, Hans Kruuk and Jane and Hugo van Lawick-Goodall. The book should find wide appeal amongst scientists, conservation practitioners and wildlife enthusiasts, especially those who visit the extremely popular Kalahari in ever-increasing numbers.

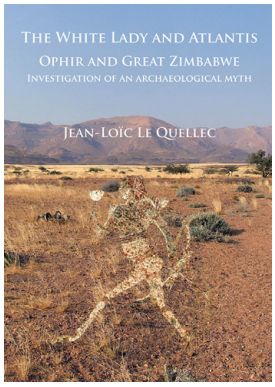
References

1. Mills G, Mills M. Hyena nights and Kalahari days. Johannesburg: Jacana; 2010.
2. Mills MGL. Kalahari hyaenas: The comparative behavioural ecology of two species. London: Unwin Hyman; 1990.

Science, stories and scholars

BOOK TITLE:
The White Lady and Atlantis, Ophir and Great Zimbabwe: Investigation of an archaeological myth

BOOK COVER:



AUTHOR:
Jean-Löic Le Quellec

ISBN:
9781784914707 (softcover)

PUBLISHER:
Archaeopress, Oxford; GBP45

PUBLISHED:
2016

REVIEWER:
Anne Solomon

AFFILIATION:
Independent Researcher, Vale of Glamorgan, Wales, United Kingdom

EMAIL:
solomon.annec@gmail.com

HOW TO CITE:
Solomon A. Science, stories and scholars. S Afr J Sci. 2017;113(11/12), Art. #a0239, 1 page. http://dx.doi.org/10.17159/sajs.2017/a0239

Scientific discoveries have always inspired writers and permeated popular culture, most obviously in science fiction. Sci-fi's precursor was the 'scientific romance' – a booming 19th and early 20th century genre. In this volume, Jean-Löic Le Quellec, a French archaeologist and specialist in Saharan rock art, examines this sub-genre, as exemplified by Conan Doyle's 1912 novel *The Lost World*, but prefigured by 'lost race', rather than 'lost world', narratives such as Rider Haggard's *King Solomon's Mines* (1885) and *She* (1886). Rather than considering the impact of how science meets stories, Le Quellec considers its inverse: 'The blurring of frontiers between science and literature, in which novelists discuss scientific questions and scholars tell stories' (p. 173). The book is profusely illustrated in colour and black and white. A small quibble is that the figures are not numbered; another is the absence of an index. However, these do not detract from the overall value of the work.

The study's starting point is the Namibian rock painting dubbed the 'White Lady of the Brandberg' that became famous after the abbé Henri Breuil (1877–1961) published a book of the same name in 1955.[1] Breuil, a French priest and archaeologist, was known as the 'father of prehistory' by his fans and the 'pope of prehistory' by his critics. After valuable work on European Palaeolithic cave art, Breuil worked for a decade in South Africa at the invitation of Jan Smuts. The 'White Lady' (which is neither white nor female) was 'discovered' in 1918 by Reinhard Maack, a German geologist, but attracted little attention until Breuil published his ideas (Chapter 1).[1,2] They were, in essence, that the 'lady' and other figures represented foreigners of Mediterranean (perhaps Cretan) or Egyptian origin.[2(p.11)] The subtext was the racist notion that indigenous peoples could not have produced such works – a prominent theme also in initial 'scientific' investigations of Great Zimbabwe (Chapter 5). As it turned out, much of Breuil's 'White Lady' interpretation was the work of his assistant, Mary Boyle, whose thinking plainly drew heavily on 'lost world' narrative themes (Chapter 3).

Le Quellec meticulously documents the surprising size and reach of this genre, discussing many of perhaps a thousand such tales, in English and French literature, Afrikaner culture and film (Chapters 8, 11, 13). The remarkably uniform themes and motifs of these stories are familiar. They feature pockets of surviving ancient peoples – Phoenicians, Egyptians, Sabaeans, Hamites, Assyrians and Romans, to name a few – encountered by intrepid European explorers in Africa or other wild places; powerful white African queens, hideous cannibals and other clichés. Le Quellec then draws out connections between the archaeological interpretations of Breuil and others (in rock art, notably Frobenius and, in the Sahara, Henri Lhote) and these dashing adventure stories. Often written as if accounts of real expeditions and experiences, some even included photos to enhance the illusion of factuality. Also discussed in some detail are the prominent motifs of lost civilisations, many with fabulous stores of gold and treasure. Ophir turns up in Great Zimbabwe and West Africa (Chapter 10), Atlantis is found in the Sahara desert (Chapter 6), while bull-jumpers from Knossos and Phoenician ships appear in African rock art.

The irony, Le Quellec suggests, is that Breuil's reading of the Namibian images arose just as the 'lost world' genre, which clearly ran in parallel with colonial ambitions and expansions, was coming to an end. Moreover, prehistorians, such as Randall-MacIver and Caton-Thompson at Great Zimbabwe, and some South African archaeologists, had persistently debunked the notion of such foreign influence in African prehistory, from 1905 onward; yet the popular appeal of these myths endured – and lingers even today.

Although the analysis of the ways in which literature dealing with Africa served colonial agendas is not new, and the work engages rather lightly with racialised, colonial era science, these are not the author's foremost concerns. The volume is an invaluably detailed overview that extends critical scrutiny beyond southern Africa and anglophone literary and research traditions. The detailed dissection of the myth of the White Lady painting is an important contribution, of relevance beyond rock art research itself. It reveals too the machinations of power in academic circles. Although some were brave enough to challenge Breuil, others were loath to cross this powerful academic figure. Even taking account of Breuil's advanced age at this point (in his seventies when he published his 'White Lady' volume), he comes across as self-aggrandising and blinkered by self-confidence; happy both to help himself to others' work, in particular that of his devoted (if deluded) assistant, Mary Boyle, and to trivialise his predecessors' contributions. Le Quellec notes Breuil's disdain for earlier recorders of the Brandberg site, although the accuracy of his own methods was hardly better. Breuil eschewed photography in favour of tracings; he blithely documented how his team would wet the paintings to enhance their visibility, outline faint images with charcoal (thus contaminating them for any future radiocarbon dating) and then 'trace' them on paper that was barely transparent (Chapter 12).

With hindsight it is easy to spot the mind-blowing absurdity of these researchers' methods and their claims to scientific factuality that Le Quellec diligently analyses, but the study has ongoing relevance as a cautionary tale for rock art and other researchers. Le Quellec poses the question: given the permeability of scientific research and popular culture, could such scientific mythopoiesis be possible today? This may be a reference to the allegedly scientific status of shamanistic readings of rock art, which have been deeply controversial, and have similarly been analysed in relation to a crossover with popular culture, namely new age spiritualities.[3]

In sum, the book is a highly informative contribution to knowledge and an accessible read for anyone interested in rock art, African history, literature or the history and sociology of knowledge. As a bonus, there are not many academic books that can regularly make the reader laugh out loud.

References

1. Breuil H (with Mary E. Boyle and E.R. Scherz). The White Lady of the Brandberg. London: Faber and Faber; 1955.
2. Breuil H. The White Lady of Brandberg, South-West Africa, her companions and her guards. S Afr Archaeol Bull. 1948;3(9):2–11. https://doi.org/10.2307/3886881
3. Kehoe AB. Shamans and religion: An anthropological exploration in critical thinking. Long Grove, IL: Waveland Press; 2000.

Chris Barnard: South Africa's fallible king of hearts

AUTHOR:
Marina Joubert[1]

AFFILIATION:
[1]Centre for Research on Evaluation, Science and Technology (CREST), Stellenbosch University, Stellenbosch, South Africa

CORRESPONDENCE TO:
Marina Joubert

EMAIL:
marinajoubert@sun.ac.za

KEYWORDS:
visible scientists; science celebrity; heart transplant; science and mass media

HOW TO CITE:
Joubert M. Chris Barnard: South Africa's fallible king of hearts. S Afr J Sci. 2017;113(11/12), Art. #a0243, 4 pages. http://dx.doi.org/10.17159/sajs.2017/a0243

On 3 December 1967, a South African surgeon stunned the world when he performed the first human heart transplant in Cape Town's Groote Schuur Hospital – a dramatic story that has been told numerous times. Here, I will focus on the central character – Chris Barnard – and his rise to global celebrity status.

Barnard started from humble beginnings. He was born on 8 November 1922 in Beaufort West and grew up in a poor family. After school he studied medicine at the University of Cape Town and obtained his MSc and PhD from the University of Minnesota. He was 45 years old when he made medical history and was swept up by a vortex of worldwide fame. His bold surgical feat triggered his rise to celebrity, but there was more to Barnard's lasting allure.[1]

Scholars who have studied the phenomenon of celebrity within science suggest that a contingent association of factors contribute to the process of celebrification.[2-5] When all these factors occur simultaneously, serendipity takes over and it appears as if an invisible hand runs a publicity campaign on behalf of the scientist.[3] Barnard's life was a case in point.

How Barnard measures up to Goodell's view of visible scientists

In the early 1970s, American sociologist Rae Goodell conducted a landmark study of visible scientists.[2] Deploying the concepts of 'visibility' and 'celebrity' synonymously,[6] Goodell classified a new type of public scientist that emerged at the time, mainly as a result of the rising power of the mass media. She shows how media criteria begin to influence science itself and notes how visible scientists use the mass media to influence public opinion and policy. Goodell describes visible scientists as revolutionaries who question established theories and advocate for change. They are strong and assertive individuals who dominate conversations and thrive at the centre of attention; they are mavericks, tilting with the establishment; they are gladiators and even gadflies in their professions and society at large.[2]

Chris Barnard was the only South African scientist identified in Goodell's study and fitted her characteristics of a visible scientist almost perfectly. He undoubtedly challenged existing norms in medicine and was boldly innovative.[7,8] Colleagues described him as intense, assertive, determined, ambitious and hungry for success, but also volatile, stubborn, ruthless, impatient, unreasonable and with a killer instinct; he was notorious for his temper, in and outside the operating theatre.[1,7]

Goodell proposes that visible scientists share five media-focused characteristics that make them attractive to journalists and heighten their conspicuousness as protagonists in the public sphere: they have a hot topic, they are controversial, they are articulate, they have a colourful image, and they enjoy credibility within science. Based on a study of Einstein's rise to fame, Missner[3] suggests a similar set of qualities that are needed for scientists to become celebrities: they make a good impression, they give quotable quotes, they make themselves available for interviews and photos, they touch the right chords with the public, and they have the right appearance and personality. Barnard's conduct and persona matched these characteristics and qualities suggested by Goodell and Missner.

Transplanting a human heart was a courageous world-first and not without controversy. Other surgeons had invested more time in preparing for this operation, but Barnard was the first to take the risk; he had the courage and heroic attitude – some would say audacity – that was needed to proceed.[9,10]

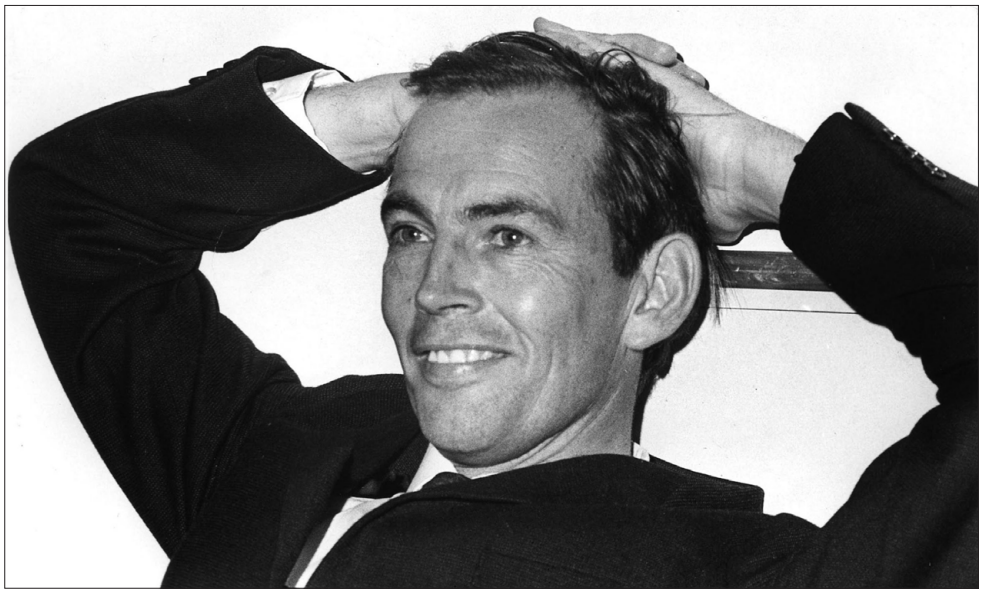


Photo: Barnardstamboom.org

Chris Barnard

When the world's media descended on Groote Schuur Hospital, Barnard's charisma and media flair ensured that journalists focused on him. His good looks and telegenic smile were media magnets, to the extent that journalists paid only scant attention to the other members of the transplant team.[1,10] Barnard himself later remarked that if he had 'been fat and bald', the media interest would have died down sooner.[1]

Amidst the avalanche of media attention, Barnard went to extraordinary lengths to accommodate journalists' demands. He happily granted interviews, posed for photos and supplied explanations. He took control of his media appearances and handled even the most difficult interviewers expertly.[1,10-14] His confidence when facing television cameras and live studio broadcasts was remarkable, given that he had no media training and hardly any media experience. (In 1967, television had not yet been introduced in South Africa.) During several publicity tours soon after the surgery, the media fell in love with him and rapidly made him world famous. In his memoirs, *One Life*[14], Barnard admitted that he found it flattering to be at the centre of so much attention.

Even before the heart transplant, Barnard enjoyed considerable academic recognition, but his breakthrough cemented his reputation and immortalised his name in medical history.[1,7-13] His article describing the surgery was published in the *South African Medical Journal* in the same month, on 30 December 1967[15], and became one of the most cited articles in the field of cardiovascular medicine[11]. In addition to more heart transplants, Barnard went on to do innovative work on prosthetic heart valves and piggy-back heart transplants, as well as the treatment of rare congenital heart defects in children. He authored or co-authored more than 200 articles in medical journals.[16]

In cases in which scientists face opposition and peer criticism, the protective effect of a solid academic reputation is well documented by scholars like Goodell[2] and Rödder[17]. It is evident that Barnard's academic credibility and public acclaim protected him from critical peers and disapproving bosses, to the extent that he became an untouchable thorn in their sides.[1]

An intimate portrayal of Barnard's vices and virtues is presented in *Barnard by Those Who Know Him*.[7] In this book, Cooper[18] describes Barnard as multi-talented, courageous and charismatic, with an unforgettable blend of vision, intelligence, charm, warmth and humour, tempered by human frailties. Lillehei[19] commends Barnard for his prodigious memory, intensity, seriousness and courage, but notes that he could provoke a rather intense dislike among some people because he was outspoken and had unconventional ideas. Frater[20] captures the contradictory aspects of Barnard as follows:

> *He was then, at once, rough-at-the-edges poor boy and charming sophisticate, democrat and tyrant, selfless healer and boorish egotist, lover and Don Juan, shrewd parvenu and naive acceptor of glitterati adulation – but, above all, surgical visionary and simply the most unforgettable character of the second generation of cardiac surgeons.*

Matching Barnard against a framework for scientific celebrity

Based on studies of celebrity, Fahy and Lewenstein[21] list six salient features that characterise celebrity scientists, all clearly illustrated by Barnard.

The first is media coverage that blurs the celebrity scientist's professional and private lives. In Barnard's case, media coverage moved on quickly from his medical achievements to his personal affairs – and there was ample material for the media to report on. He was married three times, at the ages of 25, 47 and 66. His second wife, Barbara Zoellner, daughter of a millionaire, was 19 when she married him. His third wife, Karin Setzkorn, a model, was 24 on their wedding day. Barnard had six children – two from each of his three marriages; his eldest child was born when he was 27, his youngest when he was 74. To top it all, Barnard had a reputation as a philanderer.[1] He did little to hide a series of flirtations and affairs and self-confessed to infidelities and racy sexual exploits in *The Second Life*[22] – his post-fame memoirs.

The second characteristic of celebrity scientists is that they become tradable commodities. Barnard made some money and raised funding for research from his numerous public appearances.[16] He wrote 18 books, including popular health books, autobiographies and novels, and contributed a weekly column to the *Cape Times* for many years. Later in life, he became involved in business ventures including several restaurants. He lent his name to promoting anti-ageing creams and for advertising products ranging from breakfast cereals for Kellogg's to engine oil for the Ford Motor Company.

Thirdly, the public image of a celebrity scientist is constructed around discourses of truth, reason and rationality. People were collectively fascinated by the idea that a human heart – metaphorically deeply connected to life and love – could be transplanted, but Barnard repeatedly explained that the heart was nothing more than a pump.[1] The ethical debates about the morality of removing a viable human heart were more difficult to handle and Barnard was well aware of the potential damage to his reputation around this issue.[11,23] Barnard claimed that he waited for the heart of the donor, Denise Darvall, to stop beating before removing it, but there are at least three different accounts[1] of how long it took for Darvall's heart to stop beating, all signed off by Barnard himself – in the hospital records, in the chronology of the surgery published in the *South African Medical Journal*[15] and in his autobiography *One Life*[14]. The heart transplant highlighted how Barnard benefitted from a less restrictive legal definition of clinical death in South Africa compared to that in the USA.[10,24] It would take another 10 years before the concept of brainstem death, which is fundamental to transplant surgery today, made it more acceptable to remove a beating heart.[24]

A structural relationship with the ideological tensions of their times is the fourth characteristic of scientific celebrity. This was evident in the way Barnard was co-opted to improve South Africa's image around the world at a time when the country was seen as backward because of its apartheid politics.[1,14,16] The Nationalist politicians asked Barnard to act as an ambassador for his country. A patriot at heart, Barnard mostly obliged on the world stage. At home, however, he opposed apartheid. He refused to segregate his patients according to race and inserted his political views into his public speeches and newspaper column. This led to clashes with hospital authorities and reprimands from politicians, including a warning from then State President Nic Diederichs that he could no longer protect him. The government withdrew Barnard's VIP privileges and use of private lounges at South African airports.[25]

A fifth characteristic of celebrity in science relates to tensions between scientific status and public renown. While Barnard's surgical genius and research excellence were recognised in academic circles[8,13,18,26], these were significantly overshadowed by his public acclaim when he was hailed as a hero, a heart-throb, a superman and even a miracle worker[1,27,28]. Following the heart transplant, a world of opportunities opened up to Barnard, unlike anything bestowed on a surgeon before.[1] In addition to a flood of letters, telegrams, gifts and telephone calls and a constant stream of curious onlookers in the street outside his family home,[29] he was inundated with invitations to lecture around the globe and huge numbers came to listen to him. Models, film stars, sporting heroes, royalty, religious leaders and politicians lined up to meet him. He met and dined with global icons, including Pope Paul VI, actress Sophia Loren, Princess Grace of Monaco and Princess Diana. For years, he drew huge crowds of people desperate to touch him or get hold of an autograph or photo. He was officially recognised for his pioneering work in the field of cardiac surgery by 60 countries and his numerous accolades included 15 honorary doctorates and 133 medals.[1,14,29]

Barnard's peers, however, thought that his playboy image[27] was not appropriate for a serious surgeon and they were dismayed at his excessive publicity and the way he pandered to the press. He was accused of indulging in the cult of his own personality[28] and being brazenly dedicated to his own ego, of being manipulative, reckless, and even a talented psychopath[1].

Barnard also conformed to the sixth characteristic of scientific celebrity, namely commenting on areas outside their realm of expertise.[29] He was a popular speaker on luxury ocean cruises and appeared in several advertisements. Towards the end of his career, Barnard attracted criticism when he accepted several financially attractive offers from the commercial world[1], especially when he became an advisor for a controversial research programme on so-called 'rejuvenation' therapies at a Swiss clinic and helped to promote their anti-ageing cream[29].

Barnard is also a perfect fit for the profile of a celebrity scientist suggested by Ganetz[30] based on his studies of science celebrity in the context of the Nobel Prize. The typical science celebrity, Ganetz says, is a highly educated white man who has achieved fame as the result of competition with others in the same field. The intense competition around heart transplant surgery, and Barnard's unyielding determination to win this race, is possibly one of the lesser-known aspects of his career.[9]

In 1967, several leading surgical teams around the world were poised to tackle a human heart transplant to the extent that they announced their readiness and optimism.[24] Barnard was acutely aware of these endeavours and kept an anxious eye on his rivals.[1] He was intensely concerned that he would fall behind because of the long waiting periods before new medical journals were available in South Africa and later admitted that the anxiety leading up to the heart transplant caused a painful flare-up in the arthritis that had been affecting his hands for some time.[1,14]

Barnard faced another uphill battle: he needed a patient for the procedure and for this he had to convince his superiors, in particular chief cardiologist Val Schrire, that he was ready to transplant a human heart.[1,14] Schrire eventually suggested Louis Washkansky, a critically ill 53-year-old man in the final stages of heart failure. Barnard immediately transferred Washkansky to his ward and put his surgical team at Groote Schuur on high alert. Each team member had to provide an all-hours contact number and had to be immediately available around the clock.[31] A tragic accident on 2 December 1967 delivered what Barnard was waiting for. A 25-year-old woman, Denise Darvall, was brain-dead after being hit by a speeding car, but her heart was beating strongly, thereby qualifying her to become the first human heart donor in history.

Barnard's main competitors in the race to transplant a human heart were three American surgeons – Norman Shumway, Richard Lower and Adrian Kantrowitz – as well as Donald Longmore in the UK.[9] With their extensive research into perfecting heart transplant techniques, these leading surgeons probably had little inkling of Barnard's intentions and ambitions. They must have been stumped by the news that a fairly unknown surgeon at the tip of Africa was bold enough to be the first to put the technique into practice. Following Barnard's victory, American surgeons generally downplayed Barnard's achievement, emphasising that he was merely lucky to have found the right donor at the right time.[10] The suggestion that Barnard was only able to perform his groundbreaking surgery because he stole Shumway's ideas surfaced from time to time, but Barnard dismissed this accusation as sour grapes.[1] Shumway admitted that he never liked Barnard and that it seemed cruel to have lost the race to him of all people.[9]

A love–hate relationship with the media

Many factors fuelled Barnard's rise to celebrity, but his ability to exploit the mass media played a key role. His candour, sense of humour and readiness for impromptu media briefings delighted journalists. These attributes, together with the fact that he did not use medical jargon when he talked to the press, were in stark contrast with the behaviour of his colleagues and the scientific norms of the time. In the process, Barnard probably did more than any of his peers to popularise medicine and tear down the veils of secrecy around medical procedures.[1]

In addition to extensive newspaper coverage, Barnard featured on the covers of *Life*, *Newsweek* and *Time* magazines within 3 weeks after the historic surgery. He was the focal point of several major live television broadcasts, and millions tuned in to see him on the CBS television show 'Face the Nation' on 24 December 1967 and a special edition of a BBC1 television show called 'Tomorrow's World' on 2 February 1968.

Later, however, Barnard alleged that his relationship with the media was strained at times.[32] He complained about the quality of medical journalism, the demands of the media and the invasion of his privacy.[22,32] He claimed that the skewed publicity damaged professional relationships and blamed the intrusive media coverage for harming the trust relationships between teaching hospitals and the communities they served. He also believed that the press coverage was responsible for a spate of ill-fated heart transplants that caused the operation to lose respectability as a therapeutic procedure. Barnard acknowledged that Groote Schuur Hospital was totally unprepared for the media chaos that ensued after the first heart transplant, but insisted that it would have been futile to try and moderate the publicity, as the media interest was unstoppable. He admitted, though, that the media interest resulted in research funding and institutional prestige for Groote Schuur, as well as political awareness of medical advances.[1,14]

A lasting South African legend

Barnard remained in the public eye until his death on 2 September 2001, when he suffered a severe asthma attack while alone on holiday in Cyprus. Now, in 2017, public fascination with the historic heart transplant is peaking again around the 50th anniversary of this event.

Inside the walls of Groote Schuur Hospital, a series of iconic tableaus and wax lookalikes capture the drama of the surgery and subsequent events, with the hearts of Washkansky and Darvall on public display (see http://www.heartofcapetown.co.za). Not far away, the Christiaan Barnard Memorial Hospital is a living tribute to Barnard and his medical milestone, with photos and artefacts on display in the hospital passages, and a huge plywood sculpture of an empty chest cavity suspended from the roof of the hospital foyer. Karoo Films is working on a feature film (see www.barnardfilm.com).

Reflecting on Barnard's life, Cooper[18] recalls a Native American proverb: 'A man is not dead until the last person who remembers him dies.' If this is so, Cooper says, Chris Barnard will certainly be alive for many, many years.

References

1. Logan C. Celebrity surgeon: Christiaan Barnard – A life. Johannesburg: Jonathan Ball Publishers; 2003.
2. Goodell R. The visible scientists. Boston, MA: Little, Brown and Company; 1977.
3. Missner M. Why Einstein became famous in America. Soc Stud Sci. 1985;15(2):267–291. https://doi.org/10.1177/030631285015002003
4. Turner G. Understanding celebrity. London: Sage; 2004. https://doi.org/10.4135/9781446279953
5. Fahy D. The new celebrity scientists. New York: Rowman & Littlefield; 2015.
6. Fahy D. Historical moments in public understanding of science: 1977, The Visible Scientists identifies a new scientist for the mass media age. Public Underst Sci. 2017;26(8):1019–1024. https://doi.org/10.1177/0963662517732909
7. Cooper DKC. Chris Barnard – by those who know him. Cape Town: Vlaeberg Publishers; 1992.
8. Toledo-Pereyra LH. Christiaan Barnard. J Invest Surg. 2010;23:72–78. https://doi.org/10.3109/08941939.2010.484321
9. McRae D. Every second counts: The race to transplant the first human heart. London: Simon & Schuster; 2006.
10. Nathoo A. Hearts exposed: Transplants and the media in 1960s Britain. London: Palgrave Macmillan; 2009. https://doi.org/10.1057/9780230234703
11. Brink JG, Hassoulas J. The first human heart transplant and further advances in cardiac transplantation at Groote Schuur Hospital and the University of Cape Town. Cardiovasc J Afr. 2009;20(1):31–35.
12. Hawthorne P. The transplanted heart. Johannesburg: Hugh Keartland Publishers; 1968.
13. Brink JG, Cooper DKC. Heart transplantation: The contributions of Christiaan Barnard and the University of Cape Town/Groote Schuur Hospital. World J Surg. 2005;29(8):953–961. https://doi.org/10.1007/s00268-005-0154-2

14. Barnard CN, Pepper CB. One life. Cape Town: Howard B. Timmins; 1969.

15. Barnard CN. The operation. A human cardiac transplant: An interim report of a successful operation performed at Groote Schuur Hospital, Cape Town. S Afr Med J. 1967;41(48):1271–1274.

16. Van Niekerk R, Vos H, Fouché P. The career development of Christiaan Neethling Barnard: A psychobiography. J Psychol Afr. 2015;25(5):395–402.

17. Rödder S. The ambivalence of visible scientists. In: Rödder S, Franzen M, Weingart P, editors. The sciences' media connection–public communication and its repercussions. Dordrecht: Springer; 2012. p. 155–178. https://doi.org/10.1007/978-94-007-2085-5_8

18. Cooper D. One life is enough, if well lived. In: Cooper D, editor. Chris Barnard by those who know him. Cape Town: Vlaeberg; 1992. p. 130–147.

19. Lillehei CW. Mid-life crisis? In: Cooper D, editor. Chris Barnard by those who know him. Cape Town: Vlaeberg; 1992. p. 215–227.

20. Frater R. Christiaan: The early years. In: Cooper D, editor. Chris Barnard by those who know him. Cape Town: Vlaeberg; 1992. p. 159–167.

21. Fahy D, Lewenstein B. Scientists in popular culture: The making of celebrities. In: Bucchi M, Trench B, editors. Routledge handbook of public communication of science and technology. New York: Routledge; 2014. p. 83–96.

22. Barnard C, Brewer C. The second life. Cape Town: Vlaeberg; 1993.

23. Hoffenberg R. Christiaan Barnard: His first transplants and their impact on concepts of death. BMJ. 2001;323(7327):1478–1480. https://doi.org/10.1136/bmj.323.7327.1478

24. Kantrowitz A. America's first human heart transplantation: The concept, the planning, and the furor. ASAIO J. 1998;44(4):244–252. https://doi.org/10.1097/00002480-199807000-00003

25. Molloy B. I went to the crucifixion. In: Cooper D, editor. Chris Barnard by those who know him. Cape Town: Vlaeberg; 1992. p. 240–253.

26. Cooper DKC. Christiaan Barnard and his contributions to heart transplantation. J Heart Lung Transplant. 2001;20:599–610. https://doi.org/10.1016/S1053-2498(00)00245-X

27. Malan M. Heart transplant: The story of Barnard and the 'ultimate in cardiac surgery'. Johannesburg: Voortrekker Pers; 1968.

28. Moloney G, Walker I. Messiahs, pariahs, and donors: The development of social representations of organ transplants. J Theor Soc Behav. 2000;30(2):203–227. https://doi.org/10.1111/1468-5914.00126

29. Van Niekerk R. 'n Psigobiografiese ontleding van Christiaan Neethling Barnard se loopbaanontwikkeling [A psychobiographical analysis of the career of Christiaan Neethling Barnard] [master's thesis]. Stellenbosch: Stellenbosch University; 2007. Afrikaans.

30. Ganetz H. The Nobel celebrity-scientist: Genius and personality. Celebrity Stud. 2016;7(2):234–248. https://doi.org/10.1080/19392397.2015.1088394

31. Digby A, Phillips H. At the heart of healing: Groote Schuur Hospital 1938–2008. Cape Town: Jacana Media; 2008.

32. Barnard CN. Medicine and the mass media. Am J Cardiol. 1972;30:579–580. https://doi.org/10.1016/0002-9149(72)90055-0

Promoting an environment of innovation: A university scientist's view

AUTHOR:
Brenda Wingfield[1]

AFFILIATION:
[1]NRF-DST SARChI Chair in Fungal Genomics, University of Pretoria, Pretoria, South Africa

CORRESPONDENCE TO:
Brenda Wingfield

EMAIL:
brenda.wingfield@fabi.up.ac.za

KEYWORDS:
culture of innovation; debate; networking; critical mass; financial support

HOW TO CITE:
Wingfield B. Promoting an environment of innovation: A university scientist's view. S Afr J Sci. 2017;113(11/12), Art. #a0240, 2 pages. http://dx.doi.org/10.17159/sajs.2017/a0240

Innovation is a central driving force towards achieving excellence in science. Yet I believe that we often fail to think much about what innovation really is or how to actively promote it. The Wikipedia definition – 'Innovation differs from improvement in that innovation refers to the notion of doing something different rather than doing the same thing better' – provides some insight. Nevertheless, to define innovation clearly requires us to contemplate the difference between innovation and invention. Again, Wikipedia describes this difference well: 'Innovation differs from invention in that innovation refers to the use of a better and, as a result, novel idea or method, whereas invention refers more directly to the creation of the idea or method itself.' Much for us to consider in seeking to be more effective in our research!

A special aspect of innovation is that it requires us to be in a 'special' place where we can think and experiment freely. It also requires critical mass, engagement and an exchange of ideas. Innovation seldom occurs in environments that are 'straight jacketed' by rules and bureaucracy. Unfortunately, the scientific community is subjected to increasing levels of bureaucracy – a feature of the society in which we live; a feature that is not necessarily bad, but it needs to be managed. Provision of a work place that stimulates innovation requires flexibility and agility at all levels. To illustrate this I will share a personal story. When I joined the University of Pretoria in 1998, I brought my computer (an Apple Macintosh) with me. I very soon learned that the University (at that time) did not provide support for Apple platforms. Yet I continued to use my Mac and (often frustratingly for our IT Department) I bought additional units to enable my students to succeed. When I eventually raised this matter at the highest level, I was asked why I was insisting on being difficult. The answer was actually quite simple; I needed these computers for my research and to be able to access programs that had been written only for Macs. The upshot here was that my computer should be categorised as a highly specialised piece of equipment essential for my research; the fact that it looked like a computer, operated as a computer and in fact, was a computer seemed beside the point. But this little story simply illustrates the point that research is conducted in an experimental space that needs considerable flexibility if it is going to be novel and innovative. It cannot be containerised nor packaged by administrators who commonly fail to have the background to understand this issue. To be innovative, scientists need to be able to think and experiment freely in the absence of unreasonable and unnecessarily bureaucratic restrictions. Moreover, in this regard, the best administrators of researchers must be successful researchers who understand where reasonable boundaries lie.

Innovation is what scientists should be doing every day. In many ways, it defines our training: pushing the boundaries of knowledge. Why is it then that some research programmes are more effective and innovative than others? And how can we more actively promote innovation? I certainly do not have all the answers. But after a number of decades of leading a successful research programme, I have developed opinions on some things that work and others that certainly do not. I share seven of these below.

1. Attend national and international meetings

Substantial knowledge of one's field and keeping up to date regarding the latest literature is essential. Participating in and attending talks and seminars provides a strong foundation to connect with the latest work in one's field. Here, an important aspect is attending relevant conferences globally. Ironically, in the perceptions of some organisations or research leaders, international travel is seen as a luxury; sometimes a reward for having achieved some goal. I contest this approach strongly and I view direct engagement with colleagues globally as an essential part of producing effective scientific outputs.

Understanding the value of interdisciplinary research and the need for diversity is a driving force of innovation. It is important to not limit one's reading and research horizons to one's own (often quite narrow) field. The growing – rather overwhelmingly and rapidly – body of literature can make this task seem difficult or even unattainable. I find that one of the easiest ways to deal with this challenge is to attend talks outside of my own field. In this regard, it is quite disappointing to see rather low numbers of researchers attending the many public lectures that are presented on our own campus and other campuses. Those in attendance are commonly close colleagues of the speaker. The question I commonly find myself asking is: Are we talking to only ourselves? How can we change this situation and benefit from exploring the 'bigger picture'? We learn increasingly that the world's most powerful and influential science is emerging from international collaboration, research at the intersections of disciplines and deeper levels of interdisciplinarity – from whence great innovation originates. Yet we commonly seem to be 'missing the boat' in our actions.

2. Avoid limiting interactions at meetings to friends and researchers already known to you

Talks presented at congresses are important. But it is the 'one on one' discussions at poster sessions and during breaks that really define the important outcomes of meeting attendance. This networking is essential and it is where one truly has the opportunity to discover what colleagues (sometimes competitors) are currently doing, often only to be seen in publications months later. I find it amazing to observe members of research groups huddled together at meetings rather than engaging more broadly. It might be perceived as a little dictatorial, but my students are instructed to not spend time with each other at meetings, and to rather engage with others and thus to derive the

real benefits which meetings can bring. And this is one of the strongest arguments for face to face (as opposed to web-based) meetings.

3. Debate ideas constantly

My experience has been that a 'flat' management style enhances innovation. Inordinate levels of hierarchy deter constructive discussion. 'Real' discussions require debate, often prompted by disagreement on issues being discussed. One of my greatest pleasures is finding that the best PhD students disagree with me at increasingly regular intervals as they move towards the completion of their degrees. Likewise, many of my closest research colleagues disagree with me much of the time. Sometimes they are correct, and sometimes not, but this is how we develop strong hypotheses and how we design research to test them. None of us can be correct all of the time and the best possible situation is one in which there is sufficient debate to promote the highest levels of innovation.

It is essential to have the time to talk and especially to disagree until everyone has agreed on a way forward. I set aside my Fridays to talk with students and collaborators. Some weeks we just talk about the results and the next experiment, other weeks we debate some points, and yet others we get very excited about some result or the realisation of a new hypothesis or idea. Thus, my Fridays are both exhausting and energising. Sometimes very little progress is made at these meetings and other times huge advances are made. Sometimes a whole programme can be completely turned around after a 30-min discussion, other times we try for years, unsuccessfully, to answer the same question.

4. Cultivate a stress-free research environment

Stress kills innovation. This is not to say that people who are hardworking and busy cannot be innovative; it all depends on whether they are happy and focused. One cannot be innovative when one is worrying about some or other issue. My example here is that postgraduate students need to know that they have the finances to cover their fees and reasonable board and lodging. At a certain level, this is the responsibility of the supervisor. But granting agencies can also provide support by, for example, providing scholarships on time. Likewise, administrative staff need to understand that many months of potential innovation can be destroyed in situations in which they do not understand that doing their jobs efficiently is an essential part of the innovation process. And researchers need to know that they will have the support they require in terms of dealing with their finances and budgets.

In many ways, whether one feels stressed is a state of mind. The important issue is to work out how to get to that 'mind space' in which one feels free to innovate. I do believe that this is an advantage that researchers in more developed nations have over those of us working in developing world countries. Unfortunately for South Africa, the recent protests on university campuses have not helped to provide a stress-free environment, and this situation must set back innovation.

5. Apply for financial support but do not make money your goal

Money does not drive innovation. But having sufficient financial support for research is needed to provide the space and tools required to drive innovation. What drives innovation is the perception of the space and freedom to do whatever one wishes.

6. Try to achieve critical mass

To attain an environment of innovation requires critical mass in terms of human capacity and experience. My view of such critical mass is that once we establish a technique or some technological process, we always have someone in the group able to sustain that technique and to teach it to others. In essence, we are not continuously re-inventing the proverbial wheel. This is much easier to achieve in a research institute in which there are permanent staff. In the university environment, we rely on postgraduate students and post-doctoral fellows who are, by definition, temporarily present, which presents a special challenge for maintaining critical mass.

7. Have fun

The fun and excitement of being a scientist is that one gets to do new and different things. One has the privilege of pushing the boundaries of knowledge; to go where no one has gone before. This sounds like something out of a science fiction movie, and in a way it is. Promoting innovation is complex and time consuming. I do not pretend to have all the answers, but when one *does* get it right, it is a very good place to be in.

South African carbon observations: CO_2 measurements for land, atmosphere and ocean

AUTHORS:
Gregor T. Feig[1]
Warren R. Joubert[1,2]
Azwitamisi E. Mudau[1]
Pedro M.S. Monteiro[1]

AFFILIATIONS:
[1]Natural Resources and the Environment, Council for Scientific and Industrial Research, Pretoria, South Africa
[2]Cape Point GAW Station, Climate and Environmental Research and Monitoring, South African Weather Service, Stellenbosch, South Africa

CORRESPONDENCE TO:
Gregor Feig

EMAIL:
gfeig@csir.co.za

KEYWORDS:
pCO_2; eddy covariance; carbon reporting; carbon flux

HOW TO CITE:
Feig GT, Joubert WR, Mudau AE, Monteiro PMS. South African carbon observations: CO_2 measurements for land, atmosphere and ocean. S Afr J Sci. 2017;113(11/12), Art. #a0237, 4 pages. http://dx.doi.org/10.17159/sajs.2017/a0237

Carbon dioxide plays a central role in earth's atmospheric, ocean and terrestrial systems.[1,2] About 40% of the total anthropogenic emissions since 1750 have remained in the atmosphere, with the balance being removed by the ocean and vegetation sinks.[3] Increasing atmospheric CO_2 concentrations have been well documented,[3] as have widespread impacts on human and natural systems, such as warmer surface temperatures, ocean warming and decreasing pH, loss of ice mass over the cryosphere, increasing global mean sea level, and alterations in the global hydrological cycle.[3,4] The impact of increased atmospheric concentrations of CO_2 on the biosphere includes shifting species extent, seasonal activities, migration patterns and abundances, as well as changes in species interactions.

Monitoring of atmospheric CO_2 and other greenhouse gases (GHGs) has been identified as a priority by international agencies, such as the United Nations Framework Convention on Climate Change and government departments that are interested in mitigating the effects of climate change. South Africa has made a commitment to a low carbon future as part of its role in global climate policy instruments through a national low carbon development strategy.[5,6] At the Conference of the Parties in November 2015 (COP21), high level of agreement by developed and developing countries encouraged stakeholders to urgent action to address climate change. The agreement emphasises the urgent mitigation pledges with respect to GHG emissions by 2020. As South Africa implements its White Paper on Climate Change, to stimulate a shift towards a low carbon economy, it faces a monitoring and evaluation challenge. Currently, the South African GHG emission inventory is based on fossil fuel emissions, as part of the National Atmospheric Emissions Inventory System, under the *National Air Quality Act, 2004* (*Act No. 39 of 2004*). Briefly, emissions are rarely measured directly, but rather based on proxy estimates of activity, extrapolated by an emission factor for the specific activity. There is therefore a need to independently assess the effectiveness of emissions reductions within the context of natural CO_2 fluxes. Understanding the changing driving forces of climate change and evaluation of the carbon emission reduction activities requires long-term and high-precision measurements of CO_2 gas emissions and sinks as well as their evolution.

Land can act as both a source and a sink for GHGs.[7] Currently the baseline GHG emissions from land and agriculture are thought to amount to 3.03×10^{10} kg CO_2 eq per year in South Africa. The land sector is responsible for an uptake of 2.1×10^{10} kg CO_2 eq per year while agriculture is responsible for a release of 5.06×10^{10} kg CO_2 eq per year.[7] The GHG emissions for South African industry amounted to $\sim 5.45 \times 10^{11}$ kg CO_2 eq in 2010[8,9], with approximately 79% from the energy sector – an order of magnitude larger than the emissions from agriculture[7].

Under the proposed White Paper policy, South Africa's GHG peak, plateau and decline trajectory anticipates emissions to peak at 6.1×10^{11} kg CO_2 eq between 2020 and 2025, plateau at this range for about 10 years and decline to $\sim 4.3 \times 10^{11}$ kg CO_2 eq by 2050.[6] Determining these fluxes accurately will facilitate assessment of the proposed commitments to mitigation and adaptation strategies adopted by South Africa. At present there is infrastructure deployed in South Africa for the measurement of the concentrations and fluxes of CO_2, which include observations in the atmosphere, on land and in the ocean.

Carbon dioxide observations in South Africa

Terresterial observations

A number of sites for terrestrial CO_2 observations are in place (Figure 1). These include:

- A network of cavity ring-down spectroscopy analysers for measurement of CO_2, CH_4 and H_2O concentrations. The placement of these instruments was guided by the inverse modelling work of Nickless et al.[10] These instruments have been used by the City of Cape Town for estimation of the CO_2 flux from the city.[11,12] The instrumentation has been set up around the country at the following locations:
  - The Cape Point Global Atmospheric Watch station: -34.35°, 18.48°; 172 metres above sea level (masl); operational since 1991; operated by the South African Weather Service.[13-16]
  - The Elandsfontein Air Quality Monitoring Station: -26.24°, 29.41°; 1747 masl; operational since April 2016; Eskom ambient air quality monitoring station.[17,18]
  - The Medupi Ambient Air Quality Monitoring Station: -23.74°, 27.54°; 900 masl; operational since January 2016.
- A network of eddy covariance flux towers with instruments located at the following sites:
  - Skukuza: -25.02°, 31.49°; 365 masl; savanna site in conservation area; operational since 2000; operated by the Council for Scientific and Industrial Research (CSIR).[19-21]
  - Malopeni: -23.83°, 31.21°; 385 masl; savanna site in conservation area; operational since 2008; operated by the CSIR.
  - Agincourt: -24.82°, 31.21°; 534 masl; savanna in communal area; operational since 2016; operated by the CSIR.
  - Vuwani: -23.14°, 30.43°; 629 masl; savanna site in communal area; operational since 2016; operated by the University of Venda.

- Middelburg: -31.52°, 25.01°; Karoo site in heavily grazed and lightly grazed agricultural area; operational since 2015; operated by Grootfontein Agricultural College and Stellenbosch University.
- Cathedral Peak: -28.9755°, 29.2359°; 1860 masl; operational since 2012; operated by the South African Environmental Observation Network (SAEON) Grasslands–Forests–Wetlands node.
- Welgegund: -26.56°, 26.93°; 1477 masl; grassland site under commercial agriculture; operational since 2010; operated by the North-West University.[22]

Marine observations

A number of different approaches has been adopted to address the needs of understanding and resolving the trends in Southern Ocean CO_2. One of the key gaps is observational-based estimates because of the geographical extent and remoteness of the Southern Ocean.[23] This gap is being addressed in two main ways. Firstly, by increasing the coverage and quality of global data sets through international coordinated efforts such as Surface Ocean CO_2 Atlas (SOCAT)[24] and supplementing these data with linear and non-linear empirical models and proxy variables. Secondly, by expanding the ship-based approaches with autonomous platforms such as floats[25] and gliders[23,26].

The ongoing data coverage in the Southern Ocean since 1995 has a seasonal bias for summer (Figure 2). The Southern Ocean Carbon and Climate Observatory's annual partial pressure CO_2 (pCO_2) observations programme on board the MV SA *Agulhas II* currently operates in the Southern Ocean basin annually. These pCO_2 observations seasonally characterise the drivers and variability of CO_2 fluxes in the Southern Ocean south of Africa. Moreover, these observations reduce the uncertainty of the mean annual flux of CO_2 in the Southern Ocean.[27] Reducing the uncertainty to less than 10% (or 0.1 Pg C/year) of the mean net uptake of CO_2 is critical to resolving interannual variability and trends of CO_2 flux in the Southern Ocean.[25,28]

An integrated carbon observation network which combines the current ongoing initiatives of ocean, atmosphere and terrestrial observations would provide essential information to decision-makers involved in mitigation targets and policy. In South Africa, quantitative measurement and monitoring of high-quality (climate-focused) carbon concentrations in the terrestrial, ocean and atmosphere domains already exist. Integrating these flux measurements across spatial scales and between the marine and terrestrial systems is essential.

Empirical modelling methodologies provide a method to utilise high-precision measurements of CO_2 to estimate CO_2 fluxes or to improve prior estimates of CO_2 fluxes. These methods have been successfully used in terrestrial systems including the City of Cape Town[11,12], and regional and global CO_2 emissions inventories[29,30]. This method relies on high accuracy measurements of atmospheric CO_2 (or other) concentrations to constrain a priori estimates of CO_2 fluxes derived from activity and emission factor estimates.[11,30]

Similarly, within the marine domain, empirical modelling provides an interim solution to estimate CO_2 fluxes accurately enough to estimate inter-annual and seasonal changes, as deterministic ocean models do not yet accurately depict the seasonality of CO_2. Empirical modelling utilises the relationship between in-situ CO_2 measurements and remotely sensed parameters (temperature, salinity, chlorophyll, etc.). The relationship is then applied to remotely sensed data for which there are no CO_2 measurements, to improve CO_2 data coverage. This approach has shown some promising potential in the North Atlantic where data coverage is more extensive[31,32], and has also been extended to the Southern Ocean[33]. Furthermore, the approach has more recently been refined by using artificial neural networks to highlight the importance of input parameters and self-organising maps, to illustrate the usefulness of empirical models as tools to reduce uncertainty of CO_2 estimates.[34]

The currently available CO_2 observation platforms allow the opportunity for spatial integration to provide national and metro policy management with an independent assessment capability of the effectiveness of emissions mitigation measures at local and regional (southern Africa) scales. It is necessary to maintain and expand the CO_2 observation network across ocean, terrestrial and atmospheric platforms in Southern Africa, to link the observations and modelling platforms in order to establish an observation-based CO_2 inventory for South Africa and to develop temporally relevant indicators of the state of the terrestrial, atmospheric and ocean carbon systems that are relevant and accessible to policymakers and the general public.

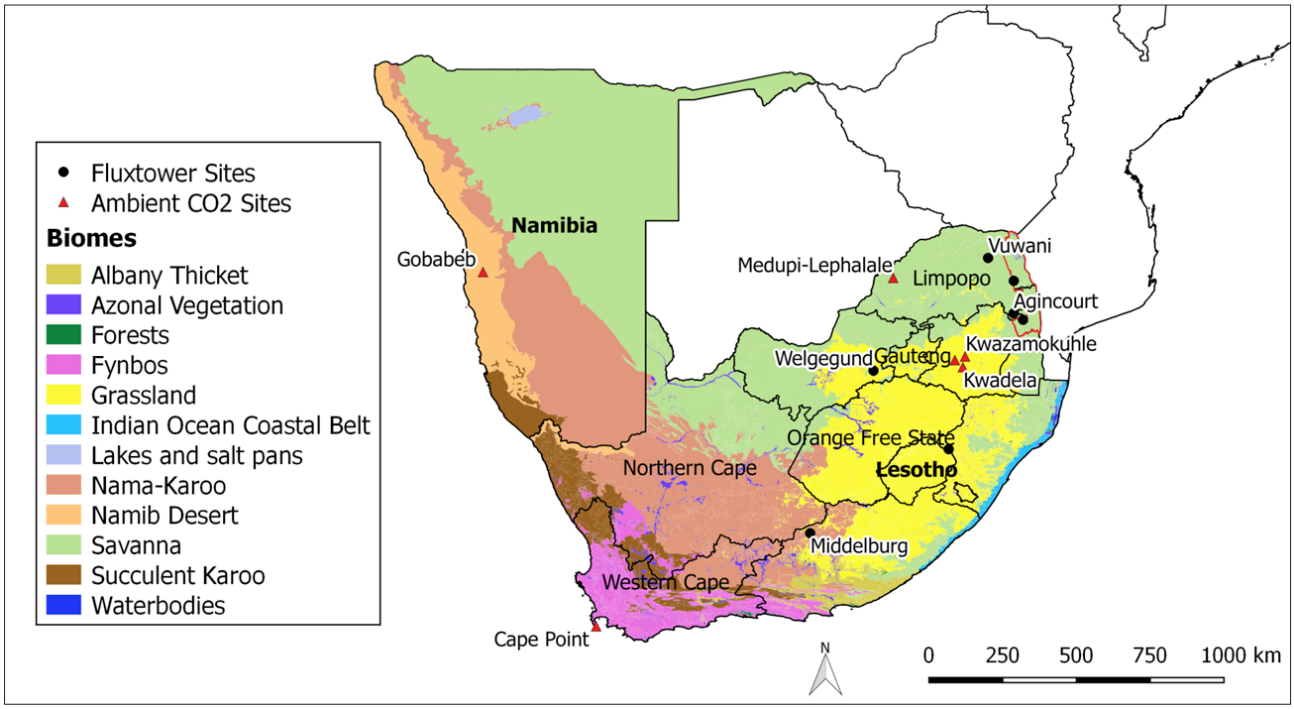


Figure 1: Sites of terrestrial CO_2 measurements in South Africa including both the flux tower locations and the CO_2 measurement sites.

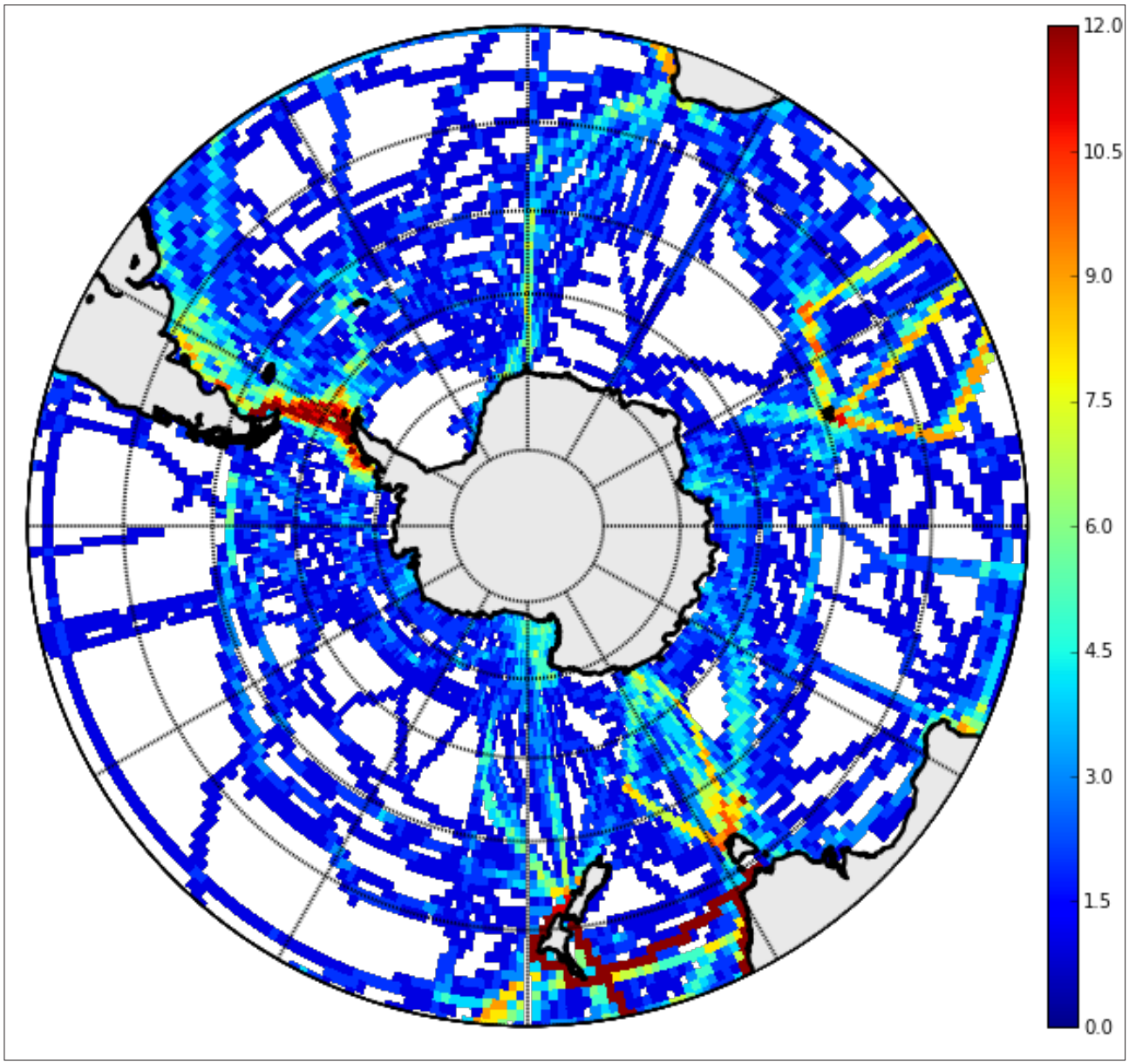


Figure 2. Gridded CO_2 observations in the Southern Ocean between 1995 and 2013 from the Surface Ocean CO_2 Atlas (SOCAT v3). The annual occupation of the seasonal cycle (in months) is shown; white space indicates no data.

References

1. Houghton RA. Balancing the global carbon budget. Annu Rev Earth Planet Sci. 2007;35:313–347. https://doi.org/10.1146/annurev.earth.35.031306.140057
2. Denman KL. Climate change, ocean processes and ocean iron fertilization. Mar Ecol Prog Ser. 2008;364:219–225. https://doi.org/10.3354/meps07542
3. Intergovernmental Panel on Climate Change (IPCC). Climate change 2014: Synthesis report: Summary for policymakers [document on the Internet]. c2014 [cited 2017 Nov 06]. Available from: http://www.ipcc.ch/pdf/assessment-report/ar5/syr/AR5_SYR_FINAL_SPM.pdf
4. Le Quéré C, Peters GP, Andres RJ, Andrew RM, Boden TA, Ciais P, et al. Global carbon budget 2013. Earth Syst Sci Data. 2014;6(1):235–263. https://doi.org/10.5194/essd-6-235-2014
5. South African National Planning Commission. National development plan (2030). Pretoria: The Presidency; 2010.
6. South African Department of Environmental Affairs (DEA). National Climate Change Response White Paper. Pretoria: DEA; 2012. p. 56.
7. Stevens L, Henri A, Van Nierop M, Van Staden E, Lodder J, Piketh SJ. Towards the deveolpment of a GHG emissions baseline for the agriculture, forestry and other land use (AFOLU) sector in South Africa. Clean Air J. 2016;26(2):34–39. http://dx.doi.org/10.17159/2410-972X/2016/v26n2a11
8. Thambiran T, Diab RD. Air quality and climate change co-benefits for the industrial sector in Durban, South Africa. Energy Policy. 2011;39(10):6658–666. http://dx.doi.org/10.1016/j.enpol.2011.08.027
9. South African Department of Environmental Affairs (DEA). GHG Inventory for South Africa. Pretoria: DEA; 2014.
10. Nickless A, Ziehn T, Rayner PJ, Scholes RJ, Engelbrecht F. Greenhouse gas network design using backward Lagrangian particle dispersion modelling – Part 2: Sensitivity analyses and South African test case. Atmos Chem Phys. 2015;15(4):2051–2069. https://doi.org/10.5194/acp-15-2051-2015
11. Nickless A, Scholes RJ, Filby E. Spatial and temporal disaggregation of anthropogenic CO_2 emissions from the City of Cape Town. S Afr J Sci. 2015;111(11–12), Art. 2014-0387, 8 pages. https://doi.org/10.17159/sajs.2015/20140387

12. Nickless A, Rayner PJ, Engelbrecht F, Brunke E, Erni B, Scholes RJ. Estimates of CO_2 fluxes over the City of Cape Town, South Africa, through Bayesian inverse modelling. Atmos Chem Phys Discuss. Preprint 2017 July. https://doi.org/10.5194/acp-2017-604

13. Nzotungicimpaye CM, Abiodun BJ, Steyn DG. Tropospheric ozone and its regional transport over Cape Town. Atmos Environ. 2014;87:228–238. Available from: http://dx.doi.org/10.1016/j.atmosenv.2014.01.063

14. Venter AD, Beukes JP, Van Zyl PG, Brunke EG, Labuschagne C, Slemr F, et al. Statistical exploration of gaseous elemental mercury (GEM) measured at Cape Point from 2007 to 2011. Atmos Chem Phys. 2015;15(18):10271–10280. https://doi.org/10.5194/acp-15-10271-2015

15. Brunke EG, Labuschagne C, Parker B, Scheel HE, Whittlestone S. Baseline air mass selection at Cape Point, South Africa: Application of 222Rn and other filter criteria to CO_2. Atmos Environ. 2004;38(33):5693–5702. https://doi.org/10.1016/j.atmosenv.2004.04.024

16. Brunke EG, Labuschagne C, Scheel HE. Trace gas variations at Cape Point, South Africa, during May 1997 following a regional biomass burning episode. Atmos Environ. 2001;35(4):777–786. https://doi.org/10.1016/S1352-2310(00)00260-0

17. Giannakaki E, Van Zyl PG, Müller D, Balis D, Komppula M. Optical and microphysical characterization of aerosol layers over South Africa by means of multi-wavelength depolarization and Raman lidar measurements. Atmos Chem Phys Discuss. 2015;15(23):35237–35276. https://doi.org/10.5194/acpd-15-35237-2015

18. Collett KS, Piketh SJ, Ross KE. An assessment of the atmospheric nitrogen budget on the South African highveld. S Afr J Sci. 2010;106(5–6), Art. #220, 9 pages. https://doi.org/10.4102/sajs.v106i5/6.220

19. Archibald SA, Kirton A, Van der Merwe MR, Scholes RJ, Williams CA, Hanan N. Drivers of inter-annual variability in net ecosystem exchange in a semi-arid savanna ecosystem, South Africa. Biogeosciences. 2009;6(2):251–266. https://doi.org/10.5194/bg-6-251-2009

20. Scholes R, Gureja N, Giannecchinni M, Dovie D, Wilson B, Davidson N, et al. The environment and vegetation of the flux measurement site near Skukuza, Kruger National Park. Koedoe. 2001;44(1):73–83. https://doi.org/10.4102/koedoe.v44i1.187

21. Feig GT, Mamtimin B, Meixner FX. Soil biogenic emissions of nitric oxide from a semi-arid savanna in South Africa. Biogeosciences. 2008;5(6):1723–1738. https://doi.org/10.5194/bgd-5-2795-2008

22. Räsänen M, Aurela M, Vakkari V, Beukes JP, Van Zyl PG, Josipovic M, et al. Carbon balance of a grazed savanna grassland ecosystem in South Africa. Biogeosciences Discuss. Preprint 2016 July. https://doi.org/10.5194/bg-2016-268

23. Waldron HN, Monteiro PMS, Swart NC. Carbon export and sequestration in the southern Benguela upwelling system: Lower and upper estimates. Ocean Sci Discuss. 2009;6(2):1173–1192. https://doi.org/10.5194/os-5-711-2009

24. Pfeil B, Olsen A, Bakker DCE, Hankin S, Koyuk H, Kozyr A, et al. A uniform, quality controlled Surface Ocean CO_2 Atlas (SOCAT). Earth Syst Sci Data. 2013;5(1):125–143. https://doi.org/10.5194/essd-5-125-2013

25. Majkut JD, Carter BR, Frölicher TL, Dufour CO, Rodgers KB, Sarmiento JL. An observing system simulation for Southern Ocean carbon dioxide uptake. Philos Trans R Soc A. 2014;372, Art. #2013-0046, 17 pages. https://doi.org/10.1098/rsta.2013.0046

26. Swart S, Chang N, Fauchereau N, Joubert W, Lucas M, Mtshali T, et al. Southern Ocean Seasonal Cycle Experiment 2012: Seasonal scale climate and carbon cycle links. S Afr J Sci. 2012;108(3–4):3–5. https://doi.org/10.4102/sajs.v108i3/4.1089

27. Thomalla SJ, Racault M-F, Swart S, Monteiro PMS. High-resolution view of the spring bloom initiation and net community production in the Subantarctic Southern Ocean using glider data. ICES J Mar Sci. 2015;72(6):1999–2020. https://doi.org/10.1093/icesjms/fsv105

28. Landschützer P, Gruber N, Bakker DCE, Schuster U, Nakaoka S, Payne MR, et al. A neural network-based estimate of the seasonal to inter-annual variability of the Atlantic Ocean carbon sink. Biogeosciences Discuss. 2013;10(5):8799–8849. https://doi.org/10.5194/bg-10-7793-2013

29. Carouge C, Peylin P, Rayner PJ, Bousquet P, Chevallier F, Ciais P. What can we learn from European continuous atmospheric CO_2 measurements to quantify regional fluxes – Part 2: Sensitivity of flux accuracy to inverse setup. Atmos Chem Phys. 2010;8(5):3119–3129. https://doi.org/10.5194/acp-10-3119-2010

30. Ogle SM, Davis K, Lauvaux T, Schuh A, Cooley D, West TO, et al. An approach for verifying biogenic greenhouse gas emissions inventories with atmospheric CO_2 concentration data. Environ Res Lett. 2015;10(3):34012. https://doi.org/10.1088/1748-9326/10/3/034012

31. Levevre N, Watson AJ, Olsen A, Rios AF, Perez FF, Johannessen T. A decrease in the sink for atmospheric CO_2 in the North Atlantic. Geophys Res Lett. 2004;31(7):2–5.

32. Chierici M, Olsen A, Johannessen T, Trinañes J, Wanninkhof R. Algorithms to estimate the carbon dioxide uptake in the northern North Atlantic using shipboard observations, satellite and ocean analysis data. Deep Res Part II Top Stud Oceanogr. 2009;56(8–10):630–639. https://doi.org/10.1016/j.dsr2.2008.12.014

33. Landschützer P, Gruber N, Bakker DCE, Schuster U. Recent variability of the global ocean carbon sink. Global Biogeochem Cycles. 2014;28(9):927–949. https://doi.org/10.1002/2014GB004853

34. Jamet C, Moulin C, Lefèvre N. Estimation of the oceanic pCO_2 in the North Atlantic from VOS lines in-situ measurements: Parameters needed to generate seasonally mean maps. Ann Geophys. 2007;25(11):2247–2257. https://doi.org/10.5194/angeo-25-2247-2007

Desalination and seawater quality at Green Point, Cape Town: A study on the effects of marine sewage outfalls

AUTHORS:
Leslie Petrik[1]
Lesley Green[2]
Adeola P. Abegunde[1]
Melissa Zackon[2]
Cecilia Y. Sanusi[1]
Jo Barnes[3]

AFFILIATIONS:
[1]Environmental and Nano Science Group, Department of Chemistry, University of the Western Cape, Cape Town, South Africa
[2]Environmental Humanities South and Department of Anthropology, School of African and Gender Studies, Anthropology and Linguistics, University of Cape Town, Cape Town, South Africa
[3]Senior Lecturer Emeritus, Division of Community Health, Stellenbosch University, Stellenbosch, South Africa

CORRESPONDENCE TO:
Lesley Green

EMAIL:
lesley.green@uct.ac.za

KEYWORDS:
microbial pollution; pharmaceuticals; perfluorinated compounds; common household chemicals; marine organisms

HOW TO CITE:
Petrik L, Green L, Abegunde AP, Zackon M, Sanusi CY, Barnes J. Desalination and seawater quality at Green Point, Cape Town: A study on the effects of marine sewage outfalls. S Afr J Sci. 2017;113(11/12), Art. #a0244, 10 pages. http://dx.doi.org/10.17159/sajs.2017/a0244

This paper presents our collection methods, laboratory protocols and findings in respect of sewage pollution affecting seawater and marine organisms in Table Bay, Cape Town, South Africa, then moves to consider their implications for the governance of urban water as well as sewage treatment and desalination. A series of seawater samples, collected from approximately 500 m to 1500 m offshore, in rock pools at low tide near Granger Bay, and at a depth under beach sand of 300–400 mm, were investigated for the presence of bacteriological load indicator organisms including *Escherichia coli* and *Enterococcus* bacteria. A second series of samples comprised limpets (*Patella vulgata*), mussels (*Mytilus galloprovincialis*), sea urchins (*Tripneustes ventricosus*), starfish (*Fromia monilis*), sea snails (*Tegula funebralis*) and seaweed (*Ulva lactuca*), collected in rock pools at low tide near Granger Bay, and sediment from wet beach sand and where the organisms were found, close to the sites of a proposed desalination plant and a number of recreational beaches. Intermittently high levels of microbial pollution were noted, and 15 pharmaceutical and common household chemicals were identified and quantified in the background seawater and bioaccumulated in marine organisms. These indicator microbes and chemicals point to the probable presence of pathogens, and literally thousands of chemicals of emerging concern in the seawater. Their bioaccumulation potential is demonstrated.

In respect of proposed desalination, the findings indicate that desalinated seawater must be subjected to treatment protocols capable of removing both bacterial loads and organic chemical compounds. The terms of reference for desalination plants must specify adequate testing and monitoring of chemical compounds as well as microorganisms in the intake and recovered water. Drinking water supplied by the proposed seawater desalination plants should be carefully tested for its toxicity.

In respect of water management, our findings suggest the need for the City of Cape Town to move to an integrated water and sewage management plan that treats urban water, including seawater, as a circulating system that is integral to the health of the City, and which excludes marine outfalls.

Background to the study

The ongoing drought in the Western Cape has led to the proposal to produce drinking water via seawater desalination plants for the City of Cape Town. The terms of reference provided in the tender documents make the assumption that the tens of millions of litres a day of untreated sewage effluent discharged into the ocean via the marine outfalls located around the Peninsula are dispersed out to sea and that intake seawater to the desalination plants will contain only inorganic salts, and not organic chemical pollutants or microorganisms.

However, kayakers, long-distance swimmers, and citizen groups like the Camps Bay Ratepayers, have claimed that untreated effluent from the marine outfalls washes back to shore in specific conditions.[1,2] Where positive independent *E. coli* counts have been demonstrated, such as those collected by public health researcher Edda Weimann[3], the City has argued that the *E. coli* results are a result of stormwater run-off.

Resolving the matter requires evidence of factors that can only have been sourced from human sewage, such as specific bacteriology and pharmacological compounds that can only have entered seawater via faecal contamination from the marine outfalls and not from surface run-off. If those compounds are present, the findings have relevance to the City's desalination plants, beach management and sewage management system.

Persistent organic pollutants include pharmaceutical and personal healthcare products such as over-the-counter and prescription drugs (antibiotics, analgesics, blood lipid regulators, natural and synthetic hormones, β-blockers, antidiabetics, antihypertensives, etc.) and household products such as soaps, detergents, disinfectants, perfumes, dental care products, skin and hair products, and surfactants, as well as these compounds' degradation products.[4-6] There is growing evidence that certain emerging contaminants could affect human and environmental health. For example, the veterinary use of diclofenac, which is also a human pharmaceutical used as an anti-inflammatory treatment, was found to be responsible for the massive decline in populations of vulture species in certain areas of Asia[7]; ethinylestradiol, one of the active ingredients in the contraceptive pill, has been associated with endocrine disruption and feminisation in fish[8]; and there is concern that long-term exposure to antibiotic pharmaceuticals and disinfectant products may be contributing to the selection of resistant bacteria with significant impacts upon human health.[9] In South Africa, Ncube et al.[10] suggested a protocol for the selection and prioritisation of contaminants in drinking water. Patterton[11] surveyed seven cities in South Africa and showed the presence of 32 compounds in drinking water, predominantly pharmaceuticals and pesticides, including carbamazepine (anticonvulsant), phenytoin (antiepileptic) and diclofenac. Osunmakinde et al.[12] compiled a priority list including the antiretroviral lamivudine, based on data collected from the health sector in South Africa. These compounds could cause far more harm than the sewage itself, such as feminisation or sterility of fish populations, cancer, growth deformities, foetal abnormalities and hormonal disturbances. These compounds may bioaccumulate in marine organisms, and thus move up the food chain to humans who eat seafood, ultimately causing the same effects. Also in South Africa, Swartz et al.[13] identified carbamazepine, sulfamethoxazole (antibiotic), triclosan (biocide), bisphenol A (plasticiser) and caffeine (stimulant) amongst others as priority pollutants for water quality assessment in water reuse.

These authors stated[13]:

> *The priority list cannot be seen as an exhaustive list as each reclaimed potable water reuse project should interrogate the relevance according to the specific area to consider whether extra chemicals might need to be added to the priority list.*

These authors also reported that many of the compounds tested for, escaped through the conventional wastewater treatment plants in trace quantities into the environment. For instance, α-ethinylestradiol, which has a recommended reference dose of 0.0015 μg/L, was present in some effluents at levels of 2–6 μg/L.

This study presents the laboratory findings from seawater, sediments as well as samples of marine organisms collected near the marine sewage outfalls in Green Point, close to the site of the proposed Granger Bay desalination plant.

Water samples: Collection methods and laboratory protocols

Seawater samples for microbiological and chemical testing were collected at 22 different points (Figure 1) near Granger Bay in the months of June, July and August 2017, together with kayakers on days on which winds and swell conditions allowed for kayak trips. In addition, seven samples were taken of water in beach sands of the intertidal swash zone at depths of approximately 300 mm. All water samples for microbiological testing were collected in bottles provided by the South African Bureau of Standards (SABS) in a sealed packet 24 h prior to collection, and samples were delivered to the SABS laboratories in Rosebank, Cape Town, within 1 h of collection and stored on ice en route. Tests were requested for *E. coli* as the indicator organism of choice for checking sewage contamination in fresh water while *Enterococcus* is more stable in seawater. At the SABS laboratories, which are accredited, samples were tested in terms of the SANS 5221 protocol for *E. coli* and SANS 7899 for *Enterococcus*. Figure 1 shows the location of sampling points in the ocean and on the shoreline and flags the hot spots of contamination above the Blue Flag limits of 250 colony-forming units per 100 millilitres (CFU/100 mL).

Microbial findings

The results of microbial tests for seawater and beach water samples are consistent with kayakers' claims that on occasion the water is a health risk. While the majority of markers were clear, there was significant variability. One sample – taken 1.7 km from shore – contained an *E. coli* count of 12 650 CFU/100 mL. This sample was collected on the edge of what kayakers identified as the sewage plume that had led to several complaints, and although the plume was visibly more dense further on, the kayakers were not willing to risk paddling into it to collect additional water samples. On the same day, a sample taken 1 km from the shore contained an *E. coli* count of 4700 CFU/100 mL.

Water collected in sand evidenced similar variability. One sample contained an *Enterococcus* count of 1460 CFU/100 mL and that on another day had a count of 7200 CFU/100 mL. The majority of microbial results were within specification, as shown on the map.

Discussion

In 2014, Edda Weimann, an endocrinologist, published a paper challenging the City's use of the Blue Flag ensign to promote its beaches.[3] Her samples, taken six times at Clifton Beach over a 4-week period in early 2013, showed that only on one day was the *E. coli* level within the Blue Flag acceptable range of below 250 CFU/100 mL and *Enterococcus* below 100 CFU/100 mL. On two separate days she found the values for *E. coli* had been in the tens of thousands, and on a further two days the values ranged in the hundreds of thousands and closer to one million (10^5–10^6). Nonetheless, on every day that she had sampled Clifton's waters, the Blue Flag had been hoisted. Weimann's findings contradict Blue Flag's criteria pertaining to water quality.[14]

Both Weimann's findings and ours suggest that predictive modelling will be more effective in managing potentially hazardous beach sewage levels than the form of water quality monitoring currently used in the City via the Blue Flag protocols, which are used to assert the health of the seawater on the basis of one or two samples taken per month.[15,16]

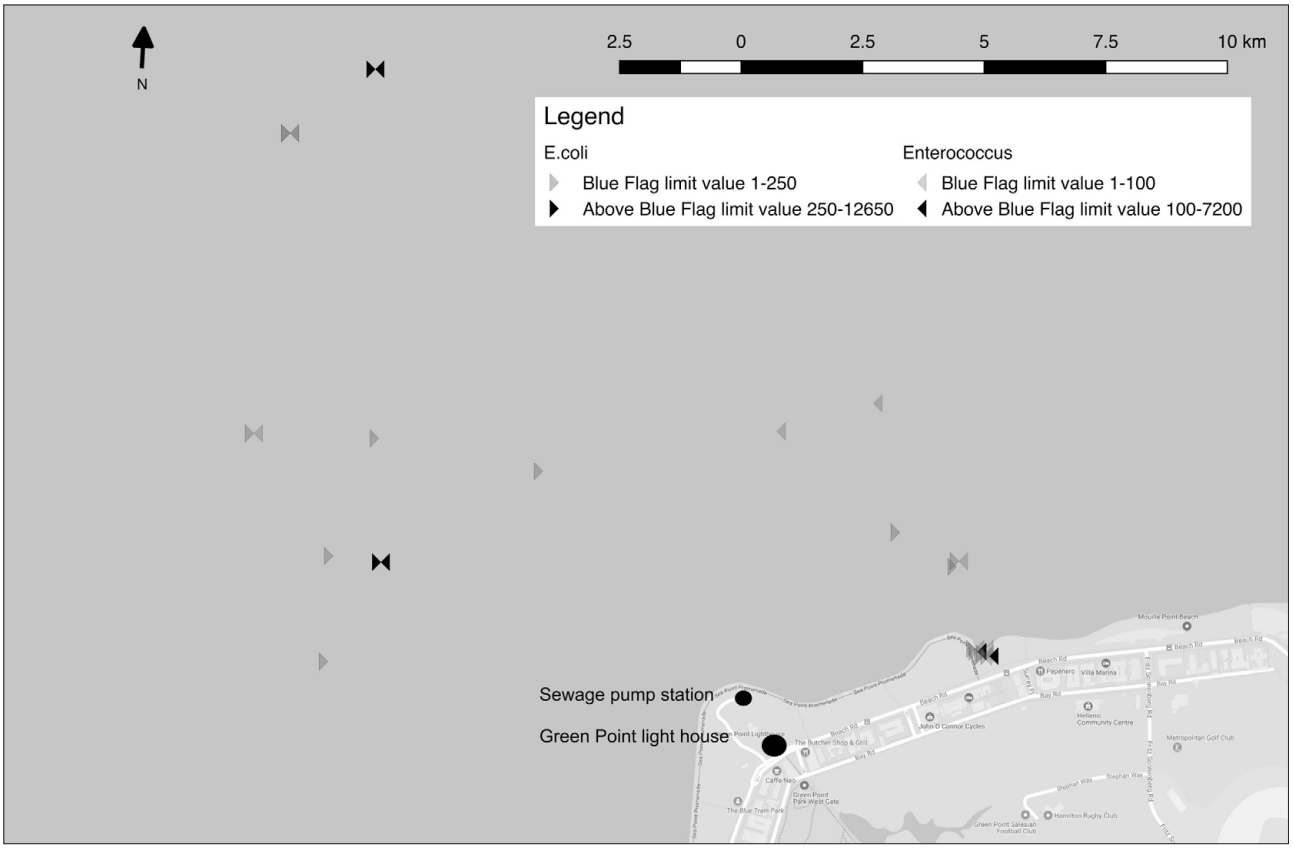


Figure 1: Location of sampling and microbial load of seawater and beach water (Granger Bay, Cape Town).

Predictive modelling is consistent with South Africa's policy for the disposal of land-derived wastewater in the sea, published in 2006.[17]

Because seawater is constantly in a state of movement we also investigated the presence in marine organisms of compounds that could only come from long-term exposure to sewage-contaminated seawater. A variety of species was collected from rock pools at low tide near Granger Bay (Figure 1).

Bioaccumulation of persistent organic pollutants in marine organisms

The selection of compounds for this study was based on their known persistence in the environment as well the availability of testing protocols and standards. The compounds tested for included perfluorinated compounds and a variety of pharmaceuticals, a cleaning agent, caffeine and bisphenol A.

Caffeine (Ca) was chosen as a broad indicator of faecal contamination. Caffeine passes from the human digestive system via faeces into the environment in unmodified form. Perfluorinated compounds are a large family of synthetic chemicals, broadly used in industrial and consumer products. They are used as industrial surfactants and surface protectors for food containers, paper, leather, carpet, fabric coating and firefighting foams because of their water and oil repelling ability. Perfluorinated compounds selected here include perfluorooctanoic acid (PFOA), perfluoroheptanoic acid (PFHpA), perfluorononanoic acid (PFNA), perfluorodecanoic acid (PFDA) and perfluoroundecanoic acid (PFUnDA). Pharmaceuticals tested for were acetaminophen (ACT), diclofenac (DSS), lamivudine (LA), phenytoin (PHE), carbamazepine (CAR) and sulfamethoxazole (SUL). The household product tested for was triclosan. Triclosan (TS) is an antibacterial and antifungal agent commonly found in household and personal cleaning products including some toothpastes. The industrial chemical tested for was bisphenol A (BPA), which is an organic synthetic compound that mimics oestrogen and is used in plastics, the lining of some food and beverage cans and thermal paper used in point-of-payment slips.

Sample collection and handling

Limpet (*Patella vulgata*), mussel (*Mytilus galloprovincialis*), sea urchin (*Tripneustes ventricosus*), starfish (*Fromia monilis*), sea snail (*Tegula funebralis*) and seaweed (*Ulva lactuca*) samples were collected from rock pools along the shoreline near Granger Bay in 2017. In 2015, samples were collected at a depth of ~30 m in the ocean close to the marine outfall diffusers. Samples were wrapped in foil and stored on ice for transportation to the laboratory. All marine organism samples were delivered to the laboratory within 1 h of collection, and stored at -20 °C at the laboratory. The samples were analysed according to the protocols below.

Analytical protocols

All sample bottles, extraction and volumetric flasks used were washed in methanol, rinsed with tap water and deionised water, then air dried.

Reagents

Methanol, acetonitrile and acetone were HPLC grade. The standards, purchased from Sigma Aldrich (Johannesburg, South Africa), were: perfluorooctanoic acid (PFOA 96%), perfluoroheptanoic acid (PFHpA 99%), perfluorononanoic acid (PFNA 97%), perfluorodecanoic acid (PFDA 98%), perfluoroundecanoic acid (PFUnDA 95%), bisphenol A (≥99%), acetaminophen (≥99%), caffeine, ibuprofen sodium salt (≥98%), diclofenac sodium salt, lamivudine (≥98%), triclosan (≥97%), phenytoin, sulfamethoxazole (≥97%), sulfisoxazole (≥99%) and acetaminophen-d_4 (≥97%). Ultrapure water was purified using a Milli-Q system (Millipore, Bedford, MA, USA).

Primary stock solutions of individual analytes were prepared in methanol at a concentration of 1000 μg/mL and appropriately diluted in methanol.

Sample preparation: Extraction and clean-up

In this study, Oasis HLB was selected over Strata X cartridge for sample extraction.[18]

Water samples

Seawater samples of 500 mL were extracted based on the method used by Valdés et al. [19], with some modifications. The extract was concentrated to 2 mL under a gentle nitrogen stream and then transferred to amber vials and centrifuged for 25 min prior to analysis.

Tissue samples

Tissue from marine organisms was freeze dried and ground into a fine powder. Approximately 10 g was weighed and placed into an extraction thimble. Surrogates (sulfisoxazole, acetaminophen-d_4) were added to each sample. The mixture was extracted with 100 mL methanol/acetone 3:1 (v/v). The extract was concentrated to 10 mL using a rotary evaporator at reduced pressure, and the sample pH was adjusted to 6 by adding 1 M NaOH or HCl so as to allow the precipitation of lipids. The extract was centrifuged at 3000 rpm for 20 min. The supernatant was transferred to glass bottles and Millipore water was added to make up to a volume of 100 mL. These aqueous extracts were further extracted and cleaned using the procedure of Valdés et al.[19] for seawater samples. The final eluate was concentrated under nitrogen and then reconstituted to 2 mL with methanol. Recovery standards were added to each sample prior to analysis.

Chromatographic conditions

The chromatographic separations were performed with the Acquity UPLC™ (Waters, Milford, MA, USA). Simultaneous determination of all the compounds of interest was achieved using an Acquity UPLC BEH C18 1.7-μm column (2.1 mm × 1000 mm) with an Acquity BEH C18 1.7-μm VanGuard™ precolumn (2.1 mm × 5 mm), supplied by Waters. The column temperature was set to 50 °C. The mobile phase consisted of a mixture of 0.02 M formic acid (solvent A) in water and acetonitrile (solvent B). Linear gradient elution of 0.35 mL/min was used starting with a mixture of 80% solvent A and 20% solvent B for 9 min. At 10 min, the acetonitrile percentage was increased linearly from 90% to 100% and was later maintained at 80% of solvent A and 20% of solvent B. A volume of 5 μL of each sample was injected into the LC/MS system. Standards and the test samples were subjected to a 12-min chromatographic run.

Mass spectrometry

The UPLC was coupled to a triple quadrupole mass spectrometer (Xevo TQ-MS), with an electrospray ionisation source. During optimisation, a multiple reaction monitoring scan mode was generated for all analytes. In addition, for maximum sensitivity, other conditions such as source temperature, capillary voltage, cone voltage, cone gas flows and desolution temperatures were standardised. This standardisation was achieved by direct injection of stock solutions with a concentration of 10 μg/mL. A capillary voltage of 3.5 kV, desolvation gas (N_2) flow of 800 L/h, source temperature of 140 °C and desolvation temperature of 400 °C were finally used. The analytical operation control and data processing were performed with Masslynx software.

Method modification, validation, quality control and calibration

The volume of each water sample used for the extraction technique was increased from 250 mL to 500 mL. To ascertain the concentration and consistency in the extraction technique for all the analytes, each extraction round was triplicated. The analytical method was validated using EU Commission Decision 2002/657/EC as a guideline. To show the applicability of the analytical method, a validation study was carried out. The validation procedure included the assessment of method linearity, specificity/selectivity, precision, recovery and calculation of the limits of detection and quantification. Six-point calibration curves were constructed (four replicates). The multi-matrix capacity of the analytical technique was checked with an identical validation study using ultrapure water and seawater. To monitor for potential contamination, blank samples of ultrapure water were extracted and analysed along with

the seawater samples and laboratory spikes. Methanol blanks were also run between samples in order to monitor for instrumental contamination and carry-over. None of the compounds of interest was detected in the ultrapure water and reagents used. Chromatographic peak area, signal noise and height were used to define and quantify the analytes of interest. Calibration standards were analysed prior to each analysis batch. The final analyte concentration was calculated as follows:

$$\text{Final analyte concentration} = \frac{\text{initial concentration x sample volume injected}}{\text{sample volume extracted}}$$

Findings: Organic pollutants in marine organisms and seawater

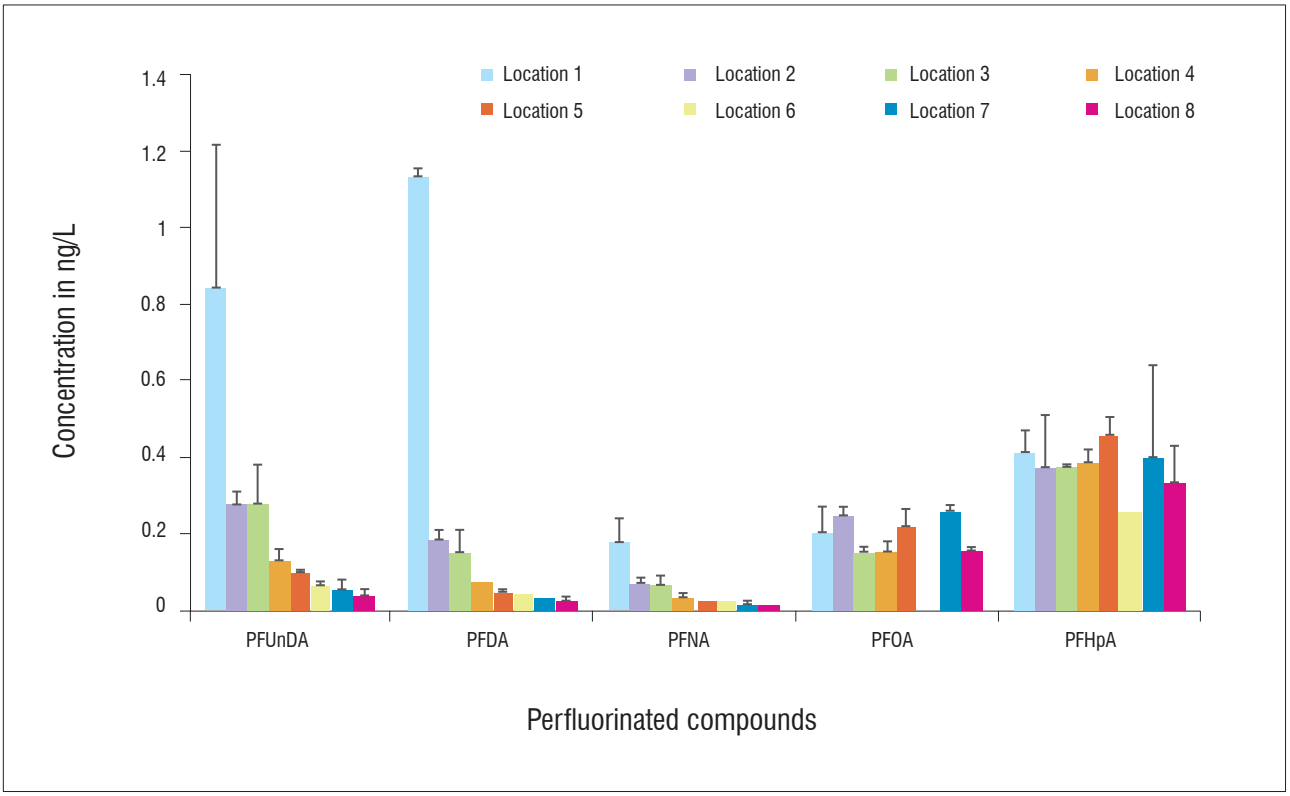
All 15 indicator chemical compounds were present in the seawater samples in trace concentrations (Figures 2 to 4) and considerably higher levels were present in limpets (*Patella vulgata*), mussels (*Mytilus galloprovincialis*), sea urchins (*Tripneustes ventricosus*), starfish (*Fromia monilis*), sea snails (*Tegula funebralis*), seaweed (*Ulva lactuca*) and sediment samples (Figures 5 to 7). The high levels of all the chemical compounds in marine organisms are evidence of bioaccumulation over time as the organisms have no way of escaping the pervasive presence of these chemicals in the seawater. The significant increase in their levels in 2017 against our findings of 2015 (Figure 8) is noteworthy. None of these compounds would normally be found in seawater and should definitely not be present in these marine organisms. With the exception of caffeine, all are manufactured substances. The finding that all 15 tested compounds were present in every organism and in the background sediments and seaweed tested, is a clear indication of faecal pollution of the shoreline, and that additional chemical substances are likely present in the seawater and thus in the marine organisms.

Discussion

Pharmacological compounds such as the analgesic and anti-inflammatory drugs acetaminophen (also known as paracetamol) and diclofenac, the anti-seizure medication phenytoin, the antibiotic sulfamethoxazole and the antiretroviral lamivudine are made to be stable and effective at low doses. They are polar, lipophilic, soluble and nonvolatile compounds.[20] For these reasons, many pharmaceutical compounds or secondary metabolites do not decompose, but survive in the environment to become persistent organic pollutants. Unknown quantities of partially metabolised drugs which may be toxic are also released in faeces and urine.[21] Their molecular sizes in the nanometre (10^{-9}) and Angstrom range (10^{-10}) make it impossible for marine organisms to exclude them. It has been widely reported that these compounds are continuously released into the environment,[22-24] and bioaccumulate in wild-caught fish populations at concentrations of nanograms per gram.[25,26] Huerta et al.[25] showed that diclofenac and carbamazepine were the most highly bioaccumulated at 18.8 ng/g in fish liver. Current regulations do not specify that they should be monitored in our water supplies or in sewage effluents (South African National Drinking Water Standard (SANS) 241: 2015), even though Patterton's study[11] demonstrated their presence in South African tap water. Moreover, it is known that disinfectants and antibiotics cause selection for resistance in the gene pool of microorganisms, ultimately making them impervious to the antibiotic or antimicrobial agents.[27]

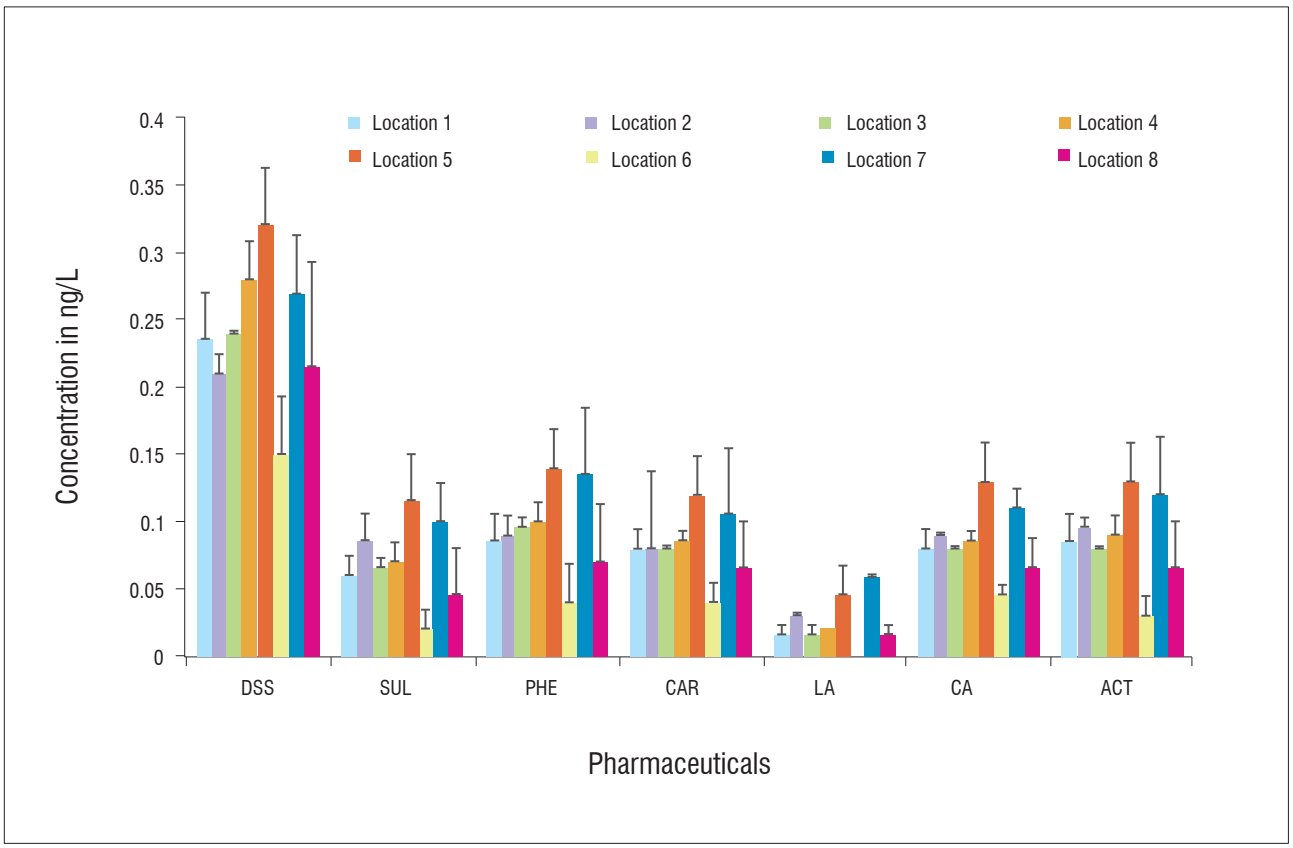
The full impact of constant, low-grade, chronic exposure to a plethora of pharmaceuticals, antibiotics and cleaning products on marine organisms, the marine food chain, and human health is not yet fully known, but their ubiquitous presence in trace levels in the desalination intake water poses a potential risk to human health.

Although some pharmaceuticals are unlikely to constitute a risk to humans as they are found in low concentrations and have a low toxicity, such as iopromide[28], other pharmaceuticals such as natural and synthetic sex hormones pose considerable risks to the aquatic environment[29].



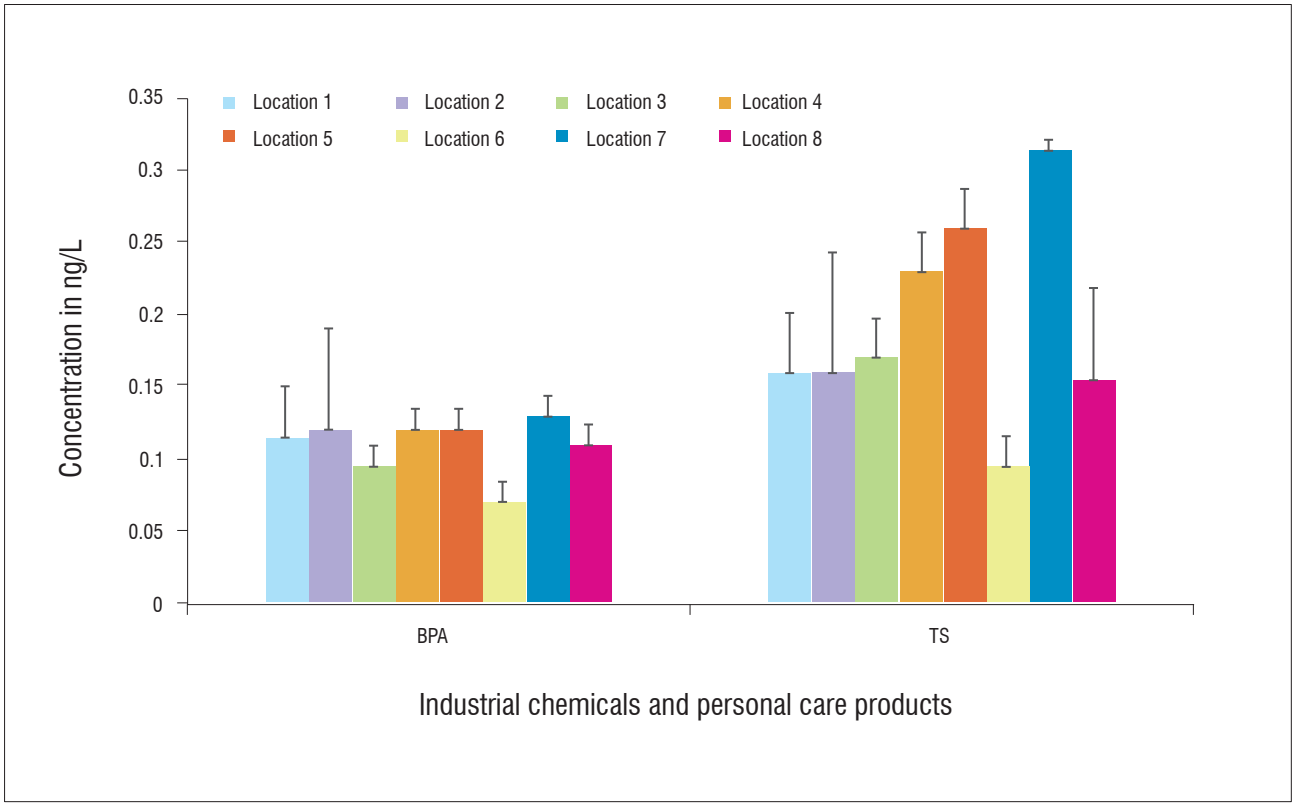
PFUnDA, perfluoroundecanoic acid; PFDA, perfluorodecanoic acid; PFNA, perfluorononanoic acid; PFOA, perfluorooctanoic acid; PFHpA, perfluoroheptanoic acid

Figure 2: Concentration of perfluorinated compounds in seawater samples collected off Granger Bay, Cape Town.



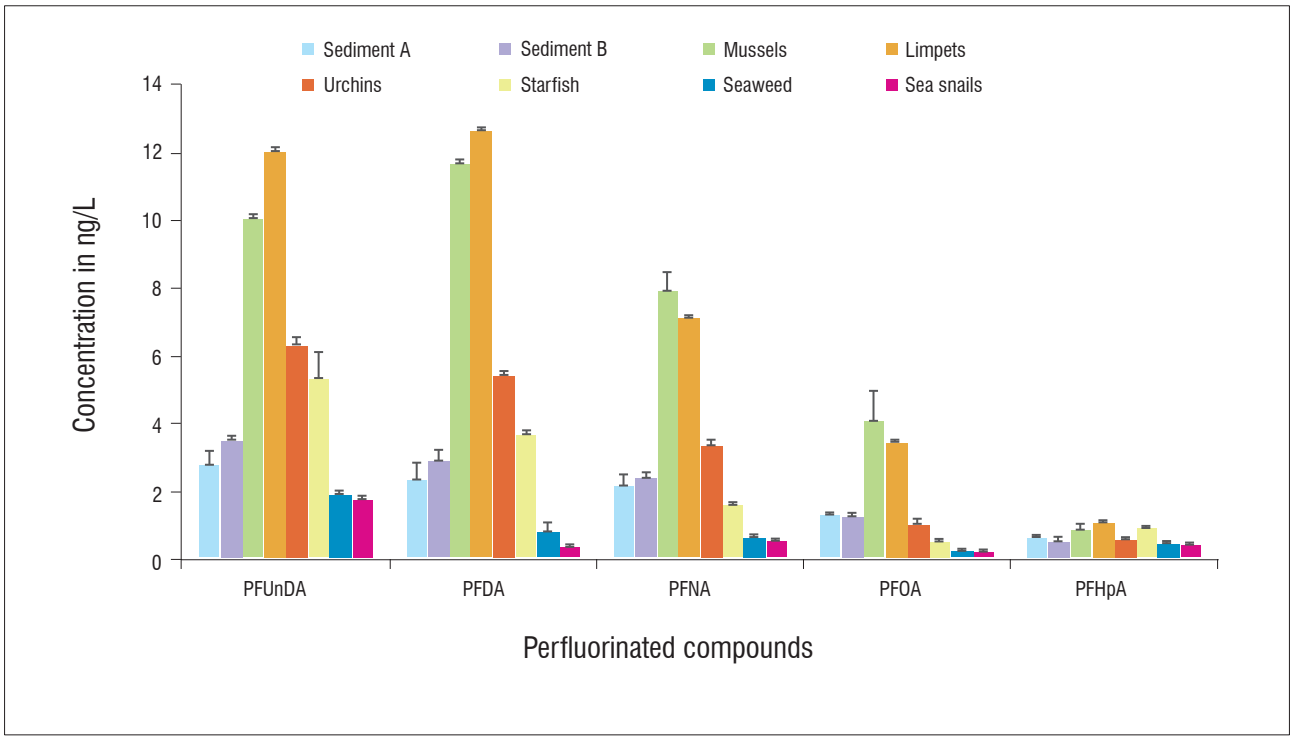
DSS, diclofenac; SUL, sulfamethoxazole; PHE, phenytoin; CAR, carbamazepine; LA, lamivudine; CA, caffeine; ACT, acetaminophen

Figure 3: Concentration of pharmaceuticals in seawater samples collected off Granger Bay, Cape Town.



BPA, bisphenol A; TS, triclosan

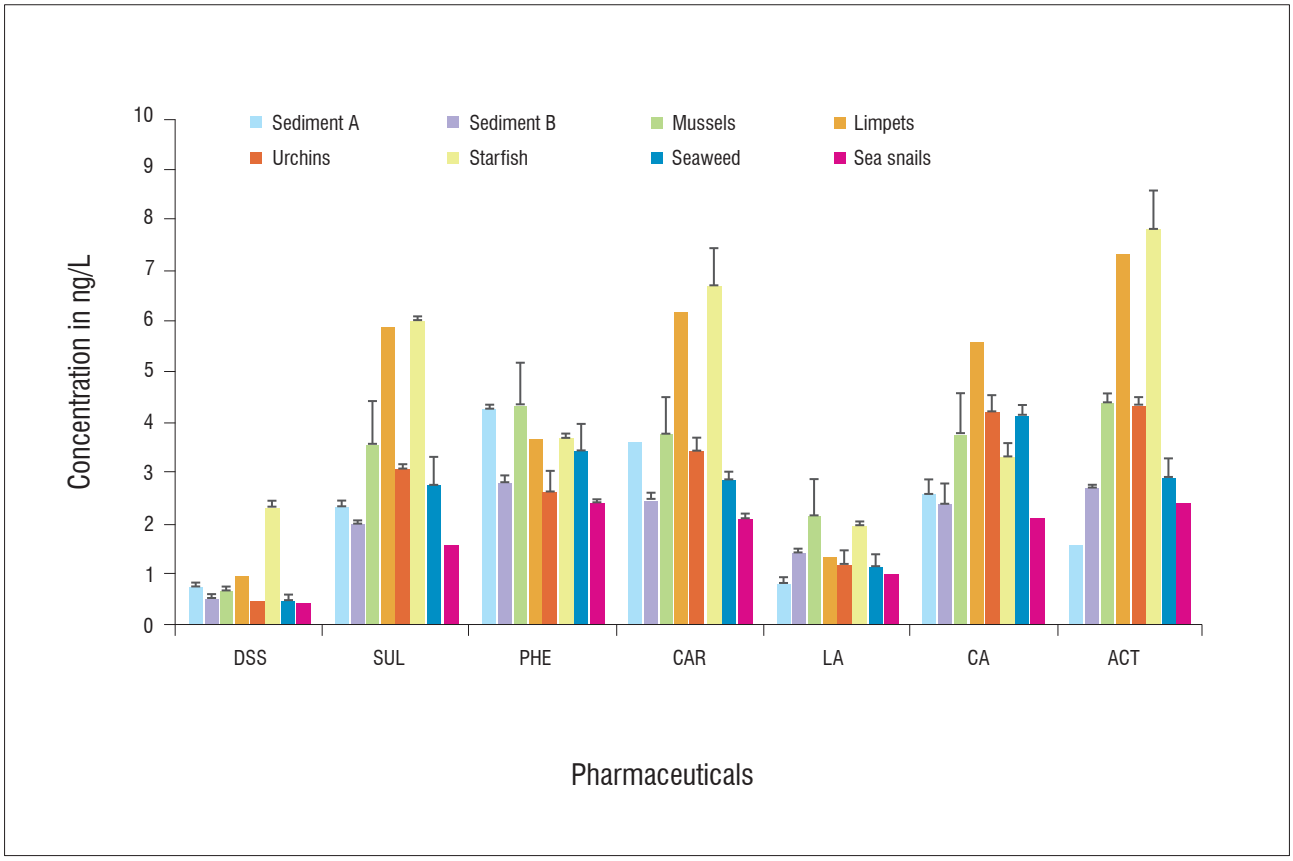
Figure 4: Concentration of industrial and household chemicals in seawater samples collected off Granger Bay, Cape Town.



PFUnDA, perfluoroundecanoic acid; PFDA, perfluorodecanoic acid; PFNA, perfluorononanoic acid; PFOA, perfluorooctanoic acid; PFHpA, perfluoroheptanoic acid

Note: Sediment A is from wet beach sand and Sediment B from where the organisms were found.

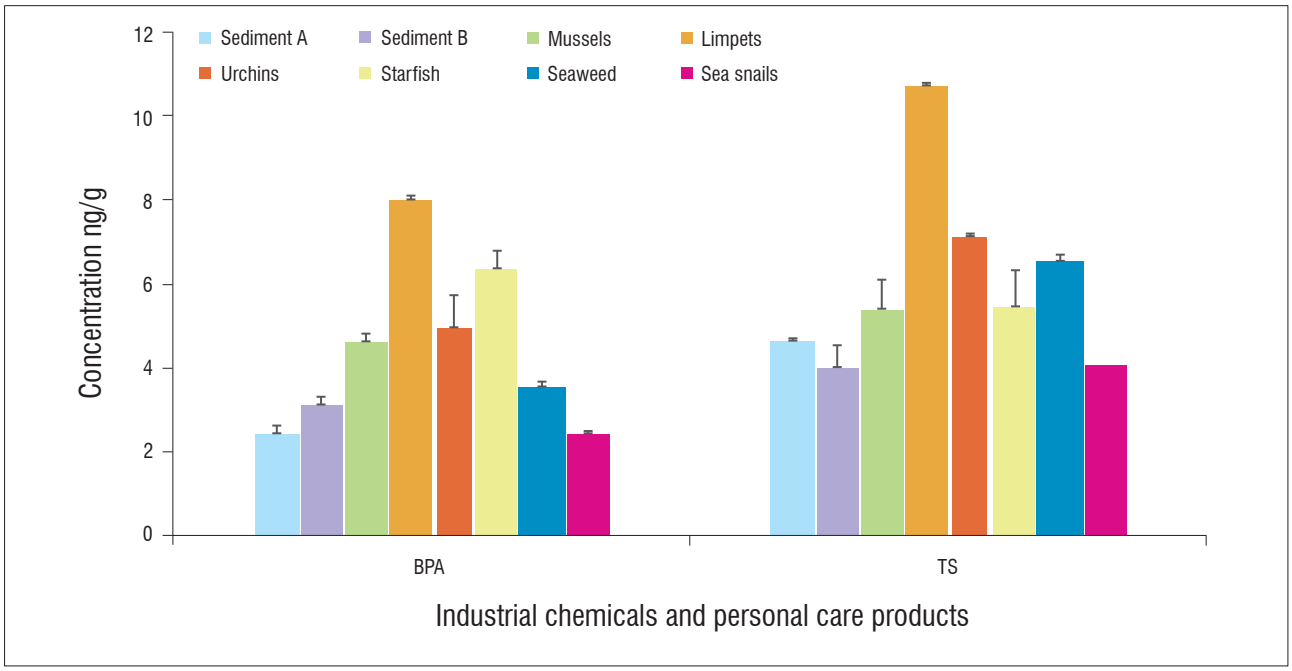
Figure 5: Concentration of perfluorinated compounds in marine organisms and sediments from the shores near Granger Bay, Cape Town.



DSS, diclofenac; SUL, sulfamethoxazole; PHE, phenytoin; CAR, carbamazepine; LA, lamivudine; CA, caffeine; ACT, acetaminophen

Note: Sediment A is from wet beach sand and Sediment B from where the organisms were found.

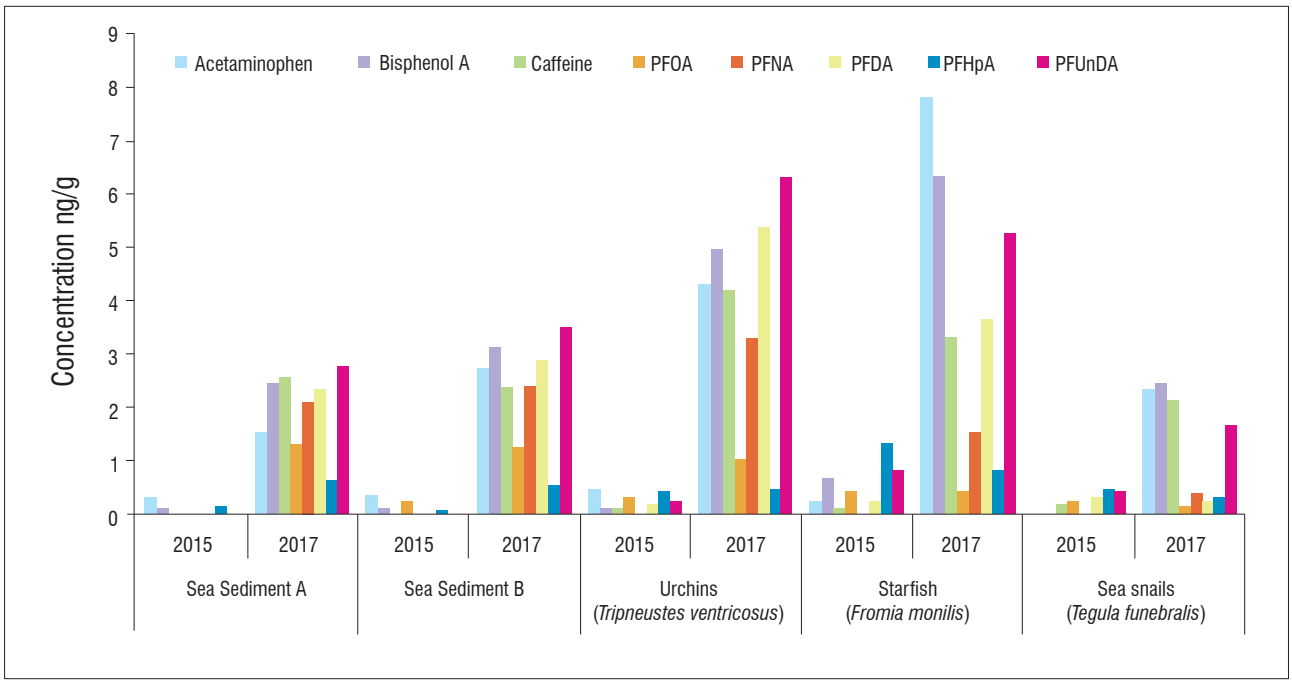
Figure 6: Concentration of pharmaceuticals in marine organisms and sediments from the shores near Granger Bay, Cape Town.



BPA, bisphenol A; TS, triclosan

Note: Sediment A is from wet beach sand and Sediment B from where the organisms were found.

Figure 7: Concentration of industrial and household chemicals in marine organisms and sediment from the shores near Granger Bay, Cape Town.



PFOA, perfluorooctanoic acid; PFNA, perfluorononanoic acid; PFDA, perfluorodecanoic acid; PFHpA, perfluoroheptanoic acid; PFUnDA, perfluoroundecanoic acid

Figure 8: The difference in the level of selected compounds in marine organisms sampled in July 2015 and July 2017.

Even where seafood accounts for only about 10% of the diet, it has been shown to be one of the main routes by which chemical contaminants find their way into human tissues[30], which in turn may be deleterious to human health. Moreover, the synergetic effects of pharmaceuticals and other compounds on living organisms are unknown.[31]

Drinking water supplies from the seawater desalination plants should be carefully tested for toxicity, which need not be costly.[13,32] Apart from the many compounds present in seawater, industrial effluents as well as hydrocarbon pollution as a result of shipping, pleasure craft and harbour effluents also impact intake water. Natural incidences such as red tide or harmful algal blooms have also been linked to marine sewage outfalls,[33-35] which could also impact the quality of intake water. As their combined effects and concentrations are mostly unknown, the precautionary principle should be followed with regard to sewage disposal into the environment. Formation of chlorine disinfection byproducts such as inorganic chloramines, organohaloginated byproducts and trichloroamines should also be monitored and removed from the recovered water.[13]

Conclusions

Implications of findings for desalination

Apart from the high microbial load being discharged into the ocean daily, the complexity and toxicity of chemicals that are being disposed into the City's sewage are imposing a growing chemical pollution risk to the nearshore coastal environment, and thus to the desalination plant's intake water. Given the diversity of contaminants shown to be ubiquitously present in the intake water in such close proximity to the marine outfall in Green Point, it is probable that the water recovered from desalination may still be contaminated with traces of complex pollutants after the reverse osmosis process, as Patterton's[11] study also showed. This probability represents a public health issue. Drinking water supplied by the seawater desalination plants should be regularly screened for its toxicity. Adequate disinfection and monitoring of the efficacy of tertiary treatment to ensure complete decomposition of harmful pharmaceuticals and other chemicals is essential to ensure that the water supplied to the City is not toxic. Screening for specific compounds is very costly but toxicity tests give rapid results.[13,32]

Even if most of the compounds were removed by the reverse osmosis step, they are not destroyed and remain in the brine retentate; returning these compounds in the brine retentate to the sea as is planned, only to be filtered indefinitely while toxic compounds build up in marine life, is a futile exercise.

In the long term, it would be technically more efficient and cost-effective to prevent the sewage from entering the ocean in the first place. Moreover, desalination intake water treatment and sewage treatment should include a tertiary stage of combined advanced oxidation, capable of fully decomposing pharmaceutical compounds. In the best performing wastewater treatment plant for potable reuse of sewage that has been studied in the Western Cape region, the treatment train was composed of a modern dual-membrane treatment process.[13] The membrane system received secondary treated wastewater from a treatment train comprised of a conventional activated sludge treatment process with an optional chemical phosphate removal after chlorination. The secondary treated wastewater entered the water recovery plant where it was treated using a sand filter, ultrafiltration membrane, reverse osmosis membranes and, finally, with advanced oxidation including ultraviolet (UV) light and hydrogen peroxide before being blended with conventionally treated water and distributed. This system currently allows direct potable reuse of sewage and is already operational in the region. This type of system may be able to provide potable water of reasonable quality from wastewater but the water from this well-operated plant was still not passing Ames mutagenicity and oestrogen mimicry tests for toxicity in our previous study.[13] With a rise in the use of chemical compounds on a daily basis, and many thousands of regulated and unregulated emerging contaminants being discharged and detected in the aquatic environment, many of which exceed the recommended reference dose (mg/kg/day) of various regulators[13], great caution is needed. Implementation of barriers, monitoring programmes and assessment programmes to eliminate or minimise these risks is essential.

Compact, new treatment systems that can treat the sewage to high standards and recover the water before discharge to the ocean can eliminate the need for desalination. Advanced oxidation systems include ozonation[36], ozone/hydrogen peroxide[37,38], ozone/UV[39], hydrogen peroxide/UV[40], UV/chlorine[41-43], UV/TiO_2[44], ultrasonic irradiation[45] or sonolysis[46,47], photocatalysis[48,49], photo fentons[50], dielectric barrier discharge[51,52] and electrochemical[53] reactions, which all work by producing short-lived but highly reactive free radicals and have been used most effectively in combined systems for the degradation or destruction of complex organic compounds in water. A thorough investigation is needed in the Western Cape on viable advanced oxidation technologies to add to the conventional treatment train of coagulation, flocculation, adsorption, precipitation, reverse osmosis, membrane bioreactors, nanofiltration and electrodialysis, recognising that the treatment of sewage and wastes is just as important to public health as the supply of fresh water is.

Implications for the City of Cape Town

The idea of sending a sewer pipeline out to sea was approved when the volumes of effluent being discharged to the ocean were relatively small, based on the incorrect assumption that 'the solution to pollution is dilution', and at a time when the variety and volume of manufactured chemicals and pharmacological compounds impacting the sewage was far lower than is the current situation.

Pipeline extension

These findings demonstrate that the assumptions behind marine sewage outfalls are incorrect and outdated. Extending the pipeline out to sea will not solve the problem, as it is clear that, under certain conditions, sewage flows back to shore in quantities that are harmful, and toxic chemicals will be released, albeit further from the shore, impacting marine life.

Predictive modelling

Until the sewer outfall is replaced, predictive modelling based on daily weather and sea conditions offers a better tool for seawater quality and beach management than sample-based monitoring. Much of the information that is required for predictive modelling is already being collected daily by ocean users who have set up WindGuru stations or similar, and who would actively participate in a citizen science project. A study consolidating such data with daily water samples is needed.

The precautionary principle

The measurable presence of indicator organisms and indicator chemicals points to the presence of pathogens and many other persistent chemicals in our ocean. The potential for their bioaccummulation is demonstrated. Because of the hazards of these compounds, the precautionary principle is highly relevant in terms of human health. Should desalination of seawater be the main option for augmenting potable water supplies, the health risks of pharmacological and chemical compound accumulation need to be quantified by daily monitoring and mitigated prior to the release of the water into the potable water reticulation system. An example of such a monitoring system is the Windhoek reclamation system. Testing the provided water to South African National Drinking Water Standard (SANS) 241: 2015 is not adequate as these compounds are not yet regulated.

The 'polluter pays principle'

In terms of the 'polluter pays principle', the costs of the chemical and pharmaceutical compound clean-up ought to be borne by the companies producing the substances. Pharmaceutical and chemical companies are among the wealthiest multinational corporations globally. While air polluters are required to ensure emissions are cleaned from the commons that is the air breathed by all, pharmaceutical companies and the chemical industry have not been contributing to the clean-up of pollutants in water systems.

Purchasing power

Retailers and consumers of pharmaceuticals and common household chemicals need to review their contribution to the growing pollution of ocean ecologies. Our individual decisions have a huge collective impact.

Politics of water, environment and sanitation

Historically, cities were made possible by the development of infrastructure to adequately manage human waste. The City of Cape Town has outgrown its current water supply and sanitation infrastructure.[54] While the City has vigorously opposed the politics of the 'poo flingers' such as Andile Lili who have dumped human waste to force the argument about improved sanitation in Khayelitsha and elsewhere, the City itself is daily depositing a volume of many Olympic-size swimming pools into the ocean. One might indeed quip that in terms of the current sewage management infrastructure, 'Je suis Andile Lili'. The convergence of sanitation activism in seaside suburbs and shack settlements in a time of drought suggests that the City's water should be understood as one

hydrological system, and therefore managed as a single ecology, not via the separation of environment, sanitation and water supply.

Acknowledgements

We gratefully acknowledge funding from the National Research Foundation (South Africa) for the Human Social Dynamics Grant (HSD141104109103 Grant 96035 titled Race and the Making of an Environmental Public) to Environmental Humanities South at the University of Cape Town. We are indebted to Tracy Fincham and Billy Fisher at Kaskazi Kayaks for guiding the sampling team and for the use of kayaks; to the SABS Rosebank for *E. coli* and *Enterococcus* results; to the laboratories at Stellenbosch University and the University of the Western Cape for the CAF results; and to Nicholas Lindenberg for assistance with the map.

References

1. Zackon M. Public protest and the production of evidence regarding Cape Town's marine sewer outfalls [honour's dissertation]. Cape Town: University of Cape Town; 2017.
2. Jackson M. Bay of sewage [video on the Internet]. c2016 [cited 2017 Nov 01]. Available from: https://www.youtube.com/watch?v=tEh5JpoH9qo
3. Weimann E. Blue Flag beaches: Bathers at risk for thalassogenic diseases. J Env Ecol. 2014;5(1), Art. #5773, 8 pages. http://dx.doi.org/10.5296/jee.v5i1.5773
4. Yu Y, Huang Q, Cui J, Zhang K. Determination of pharmaceuticals, steroid hormones and endocrine-disrupting personal care products in sewage sludge by ultra-high-performance liquid chromatography–tandem mass spectrometry. Anal Biochem. 2011;399:891–902. https://doi.org/10.1007/s00216-010-4295-2
5. Ternes TA, Siegrist H. Scrutinizing pharmaceuticals and personal care products in wastewater treatment. Environ Sci Technol. 2004;38:392A–399A. https://doi.org/10.1021/es040639t
6. Tijani JO, Fatoba OO, Babajide OO, Petrik LF. Pharmaceuticals, endocrine disruptors, personal care products, nanomaterials and perfluorinated pollutants: A review. Environ Chem Lett. 2016;14:27–49. https://doi.org/10.1007/s10311-015-0537-z
7. Oaks JL, Gilbert M, Virani MZ, Watson RT, Meteyer CU, Rideout BA, et al. Diclofenac residues as the cause of vulture population decline in Pakistan. Nature. 2004;427:630–633. https://doi.org/10.1038/nature02317
8. Lange R, Hutchinson TH, Croudace CP, Siegmund F, Schweinfurth H, Hampe P, et al. Effects of the synthetic estrogen 17 alpha-ethinylestradiol on the life cycle of the fathead minnow. Environ Toxicol Chem. 2001;20(6):1216–1227. https://doi.org/10.1002/etc.5620200610
9. Boxall ABA, Kolpin DW, Halling-Sorensen B, Tolls J. Are veterinary medicines causing environmental risks? Environ Sci Technol. 2003;37:286A–294A. https://doi.org/10.1021/es032519b
10. Ncube EJ, Voyi K, Du Preez H. Implementing a protocol for selection and prioritisation of organic contaminants in the drinking water value chain: Case study of Rand Water, South Africa. Water SA. 2012;38(4):487–503. https://doi.org/10.4314/wsa.v38i4.3
11. Patterton HG. Scoping study and research strategy development on currently known and emerging contaminants influencing drinking water quality. WRC report no. 2093/1/13. Pretoria: WRC; 2013.
12. Osunmakinde CS, Tshabalala OS, Dube S, Nindi MM. Verification and validation of analytical methods for testing the levels of PPHCPs (pharmaceutical & personal health care products) in treated drinking water and sewage. WRC report no. 2094/1/13. Pretoria: WRC; 2013.
13. Swartz CD, Genthe B, Chamier J, Petrik LF, Tijani JO, Adeleye AP, et al. Emerging contaminants in wastewater treated for direct potable re-use: The human health risk priorities in South Africa. WRC project no. K5/2369. Pretoria: WRC; 2016.
14. Blue Flag beach criteria and explanatory notes [document on the Internet]. No date [cited 2017 Jan 02]. Available from: https://static1.squarespace.com/static/55371ebde4b0e49a1e2ee9f6/t/56cc2a59859fd03dbee43223/1456220762132/Beach+Criteria+and+Explanatory+Notes.pdf
15. Olyphant GA, Whitman RL. Elements of a predictive model for determining beach closures on a real time basis. Environ Monit Assess. 2004;98(1–3):175–190.
16. Frick WE, Zepp RG. Nowcasting and forecasting concentrations of biological contaminants at beaches. Environ Sci Technol. 2008;42(13):4818–4824. https://doi.org/10.1021/es703185p
17. Taljaard S, Viljoen WAMP. Operational policy for disposal of land-derived wastewater to the marine environment of South Africa. Water SA. 2006;32(4) 527–533.
18. Adeleye AP. Perfluorinated compounds, bisphenol A and acetaminophen in selected wastewater treatment plants in and around Cape Town, South Africa [master's thesis]. Cape Town: Cape Peninsula University of Technology; 2016.
19. Valdés ME, Amé MV, Bistoni MD, Wunderlin DA. Occurrence and bioaccumulation of pharmaceuticals in a fish species inhabiting the Suquía River Basin (Córdoba, Argentina). Sci Total Environ. 2014;472:389–396. https://doi.org/10.1016/j.scitotenv.2013.10.124
20. Baker DR, Kasprzyk-Hordern B. Spatial and temporal occurrence of pharmaceuticals and illicit drugs in the aqueous environment and during wastewater treatment: New developments. Sci Total Environ. 2013;454:442–456. https://doi.org/10.1016/j.scitotenv.2013.03.043
21. Graham GG, Davies MJ, Day RO, Mohamudally A, Scott KF. The modern pharmacology of paracetamol: therapeutic actions, mechanism of action, metabolism, toxicity and recent pharmacological findings. Inflammopharmacology. 2013;21:201–232. https://doi.org/10.1007/s10787-013-0172-x
22. Boyd GR, Reemtsma H, Grimm DA, Mitra S. Pharmaceuticals and personal care products (PPCPs) in surface and treated waters of Louisiana, USA and Ontario, Canada. Sci Total Environ. 2003;311(1–3):135–149. https://doi.org/10.1016/S0048-9697(03)00138-4
23. Heberer T. Occurrence, fate, and removal of pharmaceutical residues in the aquatic environment: A review of recent research data. Toxicol Lett. 2002;131(1–2):5–17. https://doi.org/10.1016/S0378-4274(02)00041-3
24. Nunes B, Gaio AR, Carvalho F, Guilhermino L. Behaviour and biomarkers of oxidative stress in *Gambusia holbrooki* after acute exposure to widely used pharmaceuticals and a detergent. Ecotoxicol Environ Saf. 2008;71(2):341–354. https://doi.org/10.1016/j.ecoenv.2007.12.006
25. Brown JN, Paxéus N, Förlin L, Larsson JDG. Variations in bioconcentration of human pharmaceuticals from sewage effluents into fish blood plasma. Environ Toxicol Pharmacol. 2007;24:267-274. https://doi.org/10.1016/j.etap.2007.06.005
26. Huerta B, Jakimska A, Gros M, Rodríguez-Mozaz S, Barceló D. Analysis of multi-class pharmaceuticals in fish tissues by ultra-high-performance liquid chromatography tandem mass spectrometry. J Chromatogr A. 2013;1288:63-72. https://doi.org/10.1016/j.chroma.2013.03.001
27. Costanzo SD, Murby J, Bates J. Ecosystem response to antibiotics entering the aquatic environment. Mar Pollut Bull. 2005;51:218–223. https://doi.org/10.1016/j.marpolbul.2004.10.038
28. Steger-Hartmann T, Länge R, Schweinfurth H, Tschampel M, Rehmann I. Investigations into the environmental fate and effects of iopromide (ultravist), a widely used iodinated X-ray contrast medium. Water Res. 2002;36(1):266-274. https://doi.org/10.1016/S0043-1354(01)00241-X
29. Nash JP, Kime DE, Van der Ven LT, Wester PW, Brion F, Maack G, et al. Long-term exposure to environmental concentrations of the pharmaceutical ethynylestradiol causes reproductive failure in fish. Environ Health Perspec. 2004;1725–1733. https://doi.org/10.1289/ehp.7209
30. Sjödin A, Hagmar L, Klasson-Wehler E, Bjork J, Bergman A. Influence of consumption of fatty Baltic Sea fish on plasma levels of halogenated environmental contaminants in Latvian and Swedish men. Environ Health Perspec. 2000;108:35–41. https://doi.org/10.1289/ehp.001081035
31. Stackelberg PE, Furlong ET, Meyer MT, Zaugg SD, Henderson AK, Reissman DB. Persistence of pharmaceutical compounds and other organic wastewater contaminants in a conventional drinking-water-treatment plant. Sci Total Environ. 2004;329(1):99–113. https://doi.org/10.1016/j.scitotenv.2004.03.015

32. Makene VW, Tijani JO, Petrik LF, Pool EJ. Evaluation of cytotoxicity and inflammatory activity of wastewater collected from a textile factory before and after treatment by coagulation-flocculation methods. Environ Monit Assess. 2016;188:471. https://doi.org/10.1007/s10661-016-5441-x

33. Lapointe BE, Herren LW, Debortoli DD, Vogel MA. Evidence of sewage-driven eutrophication and harmful algal blooms in Florida's Indian River Lagoon. Harmful Algae. 2015;43:82–102. https://doi.org/10.1016/j.hal.2015.01.004

34. Thornber CS, Di Milla P, Nixon SW, McKinney RA. Natural and anthropogenic nitrogen uptake by bloom-forming macroalgae. Marine Poll Bull. 2008;56(2):261–269. https://doi.org/10.1016/j.marpolbul.2007.10.031

35. Paerl HW, Hall NS, Calandrino ES. Controlling harmful cyanobacterial blooms in a world experiencing anthropogenic and climatic-induced change. Sci Total Environ. 2011;409(10):1739–1745. https://doi.org/10.1016/j.scitotenv.2011.02.001

36. Qi F, Xu B, Chen Z, Ma J, Sun D, Zhang L. Efficiency and products investigations on the ozonation of 2-methylisoborneol in drinking water. Water Environ Res. 2009;81(12):2411–2419. https://doi.org/10.2175/106143009X425933

37. Pisarenko AN, Stanford BD, Yan D, Gerrity D, Snyder SA. Effects of ozone and ozone/peroxide on trace organic contaminants and NDMA in drinking water and water reuse applications. Water Res. 2012;46(2):316–326. https://doi.org/10.1016/j.watres.2011.10.021

38. Li X, Huang Y, Wang D. Efficiency and mechanism of degradation of 2-methylisoborneol (2-MIB) by O3/H2O2 in water. In: Proceedings of the 4th International Conference on Bioinformatics and Biomedical Engineering (iCBBE); 2010 June 18–20; Chengdu, China. IEEE; 2010. p. 1–4. http://dx.doi.org/10.1109/ICBBE.2010.5516282

39. Zoschke K, Dietrich N, Bornick H, Worch E. UV-based advanced oxidation processes for the treatment of odour compounds: Efficiency and by-product formation. Water Res. 2012;46(16):5365–5373. https://doi.org/10.1016/j.watres.2012.07.012

40. Rosa JM, Fileti AMF, Tambourgi EB, Santana JCC. Dyeing of cotton with reactive dyestuffs: The continuous reuse of textile wastewater effluent treated by ultraviolet/hydrogen peroxide homogeneous photocatalysis. J Cleaner Prod. 2015;90:60–65. https://doi.org/10.1016/j.jclepro.2014.11.043

41. Wang D, Bolton JR, Andrews SA, Hofmann R. UV/chlorine control of drinking water taste and odour at pilot and full-scale. Chemosphere. 2015;136:239–244. https://doi.org/10.1016/j.chemosphere.2015.05.049

42. Sarathy SR, Mohseni M. An overview of UV-based advanced oxidation processes for drinking water treatment. IUVA News. 2016;7(1):1–12.

43. Nam SW, Yoon Y, Choi DJ, Zoh KD. Degradation characteristics of metoprolol during UV/chlorination reaction and a factorial design optimization. J Hazard Mater. 2015;285:453–463. https://doi.org/10.1016/j.jhazmat.2014.11.052

44. Fotiou T, Triantis TM, Kaloudis T, O'Shea KE, Dionysiou DD, Hiskia A. Assessment of the roles of reactive oxygen species in the UV and visible light photocatalytic degradation of cyanotoxins and water taste and odor compounds using C-TiO_2. Water Res. 2016;90:52–61. https://doi.org/10.1016/j.watres.2015.12.006

45. Jiang DL, Ni GW, Zhang YM, Su YP. Algal control by low-frequency, low-power ultrasonic in eutrophic water bodies. Adv Mater Res. 2012;433–440:811–816.

46. Kohno M, Ozawa T, Niwano Y. Free radical formation from sonolysis of water in the presence of different gases. J Clin Biochem Nutr. 2011;49(2):96–101. https://doi.org/10.3164/jcbn.10-130

47. Badmus KO, Tijani J, Eze CP, Fatoba O, Petrik LF. Quantification of radicals generated in a sonicator. Anal Bioanal Chem Res. 2016;3(1):139–147.

48. Pestana CJ, Robertson PKJ, Edwards C, Wilhelm W, McKenzie C, Lawton LA. A continuous flow packed bed photocatalytic reactor for the destruction of 2-methylisoborneol and geosmin utilising pelletised TiO_2. Chem Eng J. 2014;235:293–298. https://doi.org/10.1016/j.cej.2013.09.041

49. Nyamukamba P, Tichagwa L, Mamphweli S, Petrik L. Silver/carbon codoped titanium dioxide photocatalyst for improved dye degradation under visible light. Int J Photoenergy. 2017; Art. #3079276, 9 pages. https://doi.org/10.1155/2017/3079276

50. Park J, Nama H, Choi J, Ha J, Lee S. Oxidation of geosmin and 2-methylisoborneol by the photo-Fenton process: Kinetics, degradation intermediates, and the removal of microcystin-LR and trihalomethane from Nak-Dong River water, South Korea. Chem Eng J. 2017;313:345–354. https://doi.org/10.1016/j.cej.2016.12.086

51. Mouele MES, Jimoh O, Fatoba OO, Petrik LF. Degradation of organic pollutants and microorganisms from wastewater using different dielectric barrier discharge configurations – A critical review. Environ Science Pollut R. 2015;22(23):18345–18362. https://doi.org/10.1007/s11356-015-5386-6

52. Tijani JO, Mouele MES, Tottito TC, Fatoba OO, Petrik LF. Degradation of 2-nitrophenol by dielectric barrier discharge system: the influence of carbon doped TiO_2 photocatalyst supported on stainless steel mesh. J Plasma Chem Plasma Process. 2017;37(5):1343–1373. https://doi.org/10.1007/s11090-017-9824-8

53. Li M, Xue Q, Zhang Z, Feng C, Chen N, Lei X, et al. Removal of geosmin (trans-1, 10-dimethyl-trans-9-decalol) from aqueous solution using an indirect electrochemical method. Electrochim Acta. 2010;55(23):6979–6982. https://doi.org/10.1016/j.electacta.2010.06.060

54. Robins S. The 2011 toilet wars in South Africa: Justice and transition between the exceptional and the everyday after apartheid. Dev Change. 2014;45(3):479–501.

Bacterial species from retailed poultry eggs in Tshwane, South Africa: Implication for consumers

AUTHORS:
Alexander R. Jambalang[1,2]
Elna M. Buys[3]
Francien S. Botha[1]

AFFILIATIONS:
[1]Phytomedicine Programme, Department of Paraclinical Sciences, University of Pretoria, Pretoria, South Africa
[2]Bacteriology Research Department, National Veterinary Research Institute, Vom, Nigeria
[3]Department of Food Sciences, University of Pretoria, Pretoria, South Africa

CORRESPONDENCE TO:
Francien Botha

EMAIL:
Francien.botha@up.ac.za

DATES:
Received: 05 Aug. 2016
Revised: 07 Mar. 2017
Accepted: 11 June 2017

KEYWORDS:
Enterobacteriaceae; *Salmonella*; food safety; hens' eggs; pathogenic bacteria

HOW TO CITE:
Jambalang AR, Buys EM, Botha FS. Bacterial species from retailed poultry eggs in Tshwane, South Africa: Implication for consumers. S Afr J Sci. 2017;113(11/12), Art. #2016-0232, 7 pages. http://dx.doi.org/10.17159/sajs.2017/20160232

ARTICLE INCLUDES:
× Supplementary material
× Data set

FUNDING:
None

Food safety is an important public health issue and governments across the world are intensifying their efforts to improve the quantity, quality *and* the safety of national food supplies. Bacteria, especially *Salmonella* species, present in or on chicken meat and hens' eggs in particular are the most common causes of food poisoning and the major sources of human salmonellosis. Literature reveals little information on the risk factors for salmonellae infection in Africa. The aim of this study was to determine which, if any, bacteria, especially *Salmonella* species, are present in and on hens' eggs. Representative bacterial colonies were confirmed with Gram staining and then identified using the MALDI-TOF Biotyper assay. The genera identified were *Escherichia coli* (34%), *Enterococcus faecalis* (14%), *Proteus mirabilis* (9%), *Klebshiella pneumoniae* (7%), *Salmonella* Typhimurium (6%), *Enterobacter cloacae* (1%), *Stenotrophomonas maltophilia* (0.6%), *Salmonella* Dublin (0.6%) and *Salmonella* Braenderup (0.2%). Raw hens' eggs and products containing raw hens' eggs may contain pathogenic bacteria, thereby exposing a large number of consumers to the risk of contracting food poisoning when undercooked or uncooked hens' eggs are consumed.

Significance:

- Enterobacteriaceae counts are used as an indicator to evaluate the hygienic quality of food.
- The presence of *Salmonella* species and other Enterobacteriaceae in raw hens' eggs poses a health risk to consumers.
- Any product in which raw eggs are used must be provided with a conspicuous label stating that it may contain pathogenic bacteria.

Introduction

Enterobacteriaceae is a family of Gram-negative, facultative anaerobic rod-shaped bacteria. The genera include plant and human pathogens like *Escherichia, Klebshiella, Salmonella, Shigella* and *Yersinia*.[1,2] *Escherichia coli* counts and the presence of coliforms are used as indicators to evaluate the 'hygienic' quality of raw foods.[2,3]

Microbial contamination of egg shells is of increasing concern to farmers and consumers of hens' eggs and poultry products in general.[4] In recent years, the 'farm to fork' approach to food safety has received considerable attention as a more complete method of ensuring food safety. *Salmonella* Enteritidis can be transmitted via trans-ovarian route or via faecal contamination through shell penetration to intact hens' eggs which can cause salmonellosis in consumers.[5] Food safety therefore becomes everybody's responsibility and not just the processors'.[6-8] Hens' egg farms or plants experience huge challenges in maintaining good hygienic conditions because of the high concentration of hens on the premises.[4] Fomites, flies, dust, faeces and rodents serve as a vehicle for contamination, re-contamination or cross-contamination during collection, washing, sorting, transportation and packaging of eggs.[2,8] Although great care is taken to maintain shell integrity, some eggs crack during transportation and the contents leak onto charts, providing good substrate for bacterial and fungal growth.[4]

Previous studies have found eggs and egg-processing environments to be contaminated with large numbers of Enterobacteriaceae and aerobic microorganisms. Identified isolates in eggs and on shells included *Citrobacter youngae, Enterobacter cloacae, Enterobacter sakazakii, Escherichia coli, Flavimonas oryzihabitans, Klebshiella pneumoniae, Kluyvera* spp., *Pantoea* spp., *Proteus* spp., *Providencia* spp., *Rahnella aquatilis, Salmonella* spp., *Serratia* spp., *Shigella* spp., *Xanthomonas maltophilia* and *Yersinia* spp.[9-12] Many members of this family are human pathogens, some are spoilage organisms, and others – like *Escherichia* and *Proteus* – cause various types of egg rot.[9-12]

Salmonellosis is a zoonotic food-borne bacterial disease that poses a major threat to public health and causes economic losses.[13,14] Pathogenic *Salmonella* survive in water, soil and faeces, and contaminated food like raw eggs is one of the major sources of salmonellosis in humans.[15] Treatment of salmonellosis has been complicated with the emergence of multidrug-resistant phenotypes among the *Salmonella* serotypes.[16-18] In view of the annual estimated 1.3 billion human infections globally as a result of salmonellosis and socio-economic losses of about USD1.1 billion in the USA alone[19], control measures should be put in place in order to safeguard life and contain the spread of the disease.

Literature on the distribution of different types of bacteria on hens' eggshells in South Africa revealed limited information. Experiments and research conducted elsewhere in the world have evaluated eggshell microbial populations by simulating contamination under laboratory conditions.[6-8,20]

Bacterial isolation and identification remains the gold standard to determine the presence of bacteria in food. However, a culture-based approach is not optimal because it is time consuming and not always specific.[21] Molecular techniques, on the other hand, produce rapid, automated results with high sensitivity and specificity; however, they lack the ability to isolate the organism – which is the gold standard in microbial identification.[22] Therefore,

in this study, a combination of a microbiological culture-based method (for the isolation and preliminary identification) and a rapid molecular technique (Matrix-assisted Laser Desorption Ionisation Time of Flight or MALDI-TOF) was used in the confirmatory identification (respectively) of selected bacterial species in retailed hens' eggs.

The aim of this study was to determine which, if any, bacterial species, especially *Salmonella* species, are present in and on commercial hens' eggs in the Tshwane district, Gauteng Province, South Africa.

Materials

Unwashed hens' eggs of different egg brands were randomly purchased from retail outlets in Tshwane. The egg samples were transported in ice-boxes to the Phytomedicine Laboratory, Department of Paraclinical Sciences, Faculty of Veterinary Science, Onderstepoort, South Africa and tested on the day of arrival. The major public transport routes running from north to south and east to west of Tshwane were used as sample collection sites; this choice was because a high number of commercial activities usually take place along such routes and the chances of having a similar distribution of egg retail outlets along these routes was high. A total of 468 eggs representing 13 egg brands were purchased. To protect the brands' identity, codes AJ01–AJ13 were used for brands, and samples were identified as samples AJ01–AJ468.

Methods

Microbiological sampling

All the experimental work was carried out in a Class II Biological Safety Cabinet (ESCO, Singapore) and sterile hand gloves were used during the experiments to minimise contamination. The eggs were removed from their boxes and placed with the pointed end facing down on a plastic egg crate that was sterilised with 70% ethanol (Sigma Aldrich, Johannesburg, South Africa); each egg was individually labelled.

Shell

A plastic template sterilised with 70% ethanol was used to mark out a specific area (20 mm x 20 mm) on the egg shell (using a sterilised pencil), in order to standardise the area chosen and thereby prevent biased sampling.

A swab was taken from the marked area of each egg shell using a sterile cotton swab that had been dipped in buffered peptone water (BPW) and placed into a sterile, 10-mL, screw-capped bijou bottle containing 9 mL BPW (Selecta-MEDIA, Johannesburg, South Africa).

Thereafter, the eggs were sprayed with 70% ethanol in order to disinfect the egg shell and prevent contamination of the egg content. The eggs were allowed to dry for 10 min.

Yolk and albumin

Using sterile scissors and thumb forceps, each egg was cracked open at the air sac end in order to avoid spillage. The albumin was aseptically separated from the yolk by gently decanting the albumin into sterile, wide-mouth, 30-mL, screw-capped, plastic centrifuge tubes. A sterile, 5-mL, single-channel pipette (GILSON, Villiers-le-Bel, France) was used to gently pipette out the remaining albumin, leaving behind the egg yolk. The egg yolk was carefully poured into sterile, 30-mL, screw-capped, plastic centrifuge tubes. The samples were homogenised according to the method previously described.[10]

Pre-enrichment for presumptive bacterial species

After homogenisation, 1 mL albumin and 1 mL yolk were separately pipetted into sterile bijou bottles and pre-enriched in 9 mL BPW and incubated at 37 °C for 24 h.

Selective enrichment for *Salmonella*

After pre-enrichment, 1 mL of the pre-enriched broth was used to inoculate 9 mL of Muller–Kaufmann Tetrathionate Broth (Selecta-MEDIA) for *Salmonella* selective enrichment and the broth mixture was incubated at 37 °C for 24 h. The remaining pre-enriched BPW was used for the isolation of other bacterial species on several agars like Xylose Lysine Deoxycholate agar (XLD), McConkey agar and Nutrient agar (Sigma-Aldrich). XLD agar is both a selective and differential medium for the isolation, cultivation and differentiation of *Salmonella* and *Shigella* species in particular, and most Gram-negative enteric microorganisms. Colonies of the different genera were distinguished based on morphology of the bacteria on different agar, change in colour of the agar and Gram stain reaction.

Isolation of *Salmonella* on selective agar plates

Using a calibrated inoculating loop (Sigma-Aldrich), 10 μL of the tetrathionate broth was streaked on XLD and McConkey agars (MERCK, Darmstadt, Germany) and incubated at 37 °C for 24 h. This procedure was also repeated on XLD, McConkey and nutrient agars for the isolation of bacteria other than *Salmonella* that may have been present in the enriched broth. Isolated colonies were purified and placed on ceramic beads in microbank cryoprotective media (Pro-Lab, Texas, USA) and stored at -80 °C until further analyses were performed.

MALDI-TOF assay

After the preliminary bacterial isolation and identification, all *Salmonella* and other representative presumptive bacterial isolates were subjected to the MALDI-TOF assay – a protein fingerprinting technique for confirmatory identification and biotyping to species level. The MALDI Biotyper System identifies microorganisms using MALDI-TOF mass spectrometry to measure highly abundant proteins that are found in all microorganisms. The characteristic patterns (or 'fingerprints') of these highly abundant proteins are used to reliably and accurately identify a particular microorganism by matching the respective pattern with an extensive database to determine the identity of the microorganism.[23] After the acquisition of the spectral data had been completed, a run results report was generated. The resultant report for each sample shows the best match along with the respective matching score.

The MALDI-TOF assay was done in the Department of Microbiology and Plant Pathology, University of Pretoria, using a modified method previously described by Mellmann et al.[24] The method is briefly described here.

A bacterial colony was added into an Eppendorf tube (Oxoid, Basingstoke, England) containing 300 μL deionised water; the contents of the tube were vortexed and 900 μL electrophoresis-grade ethanol (Sigma-Aldrich) was added to the bacteria-water mixture and the tube was centrifuged at 15 000 *g* for 2 min. Then 10 μL of 70% formic acid (Sigma-Aldrich) was added to the bacterial pellet and auto vortexed; 10 μL acetonitrile (Sigma-Aldrich) was also added to this mixture and the mixture was then centrifuged at 15 000 *g* for 2 min. A volume of 1 mL of supernatant containing the bacterial extract was spotted on a MALDI-TOF steel target plate and allowed to dry at room temperature. The material was next overlaid with 1 μL of a saturated solution of α-cyano-4-hydroxycinnamic acid in 50% acetonitrile, 2.5% trifluoroacetic acid (Sigma-Aldrich) within 1 h and allowed to dry at room temperature. The steel target plate was inserted into a Bruker Microflex Daltoniks MALDI Biotyper (Bruker Microflex MALDI Biotyper, Bremen, Germany) and the results were read. For bacterial identification, the spectrometer was set at linear positive mode, 60 Hz laser frequency, 20 kV acceleration voltage, 16.7 kV IS2 voltage, 170 ns extraction delay and 2000–20 137 m/z range. MALDI-TOF Biotyper 3.0 Real Time Classification by Bruker Daltoniks was used to analyse spectra for the newly investigated bacteria against the default reference settings of the Biotyper databank directly.[24] It took about 3 h to identify 96 samples from preparation to species identification and approximately 10 min for a single sample.[24]

Dendrograms of identified bacteria were generated from repeat profiles and calculated based on simple matching similarity coefficient and complete linkage from the generated and stored data using Pearson product moment correlation and gel view (BioNumerics 7.5). Ward's clustering method was used, which is a hierarchical agglomerative method whose main objective is to create clusters that give minimum increase in the total within group error sum of squares.[25]

Results

A total number of 13 different egg brands and 468 egg samples were analysed. The results indicate that 73% of the egg samples had bacterial contamination distributed on the shell, albumin and yolk as follows: *Escherichia coli* 159 (34%), *Enterococcus faecalis* 66 (14%), *Proteus mirabilis* 42 (9%), *Klebshiella pneumoniae* 33 (7%), *Salmonella* serotype Typhimurium 28 (6%), *Enterobacter cloacae* 5 (1%), *Stenotrophomonas maltophilia* 3 (0.6%), *Salmonella* serotype Dublin 3 (0.6%) and *Salmonella* serotype Braenderup 1 (0.2%). (Figure 1). However, 128 (27%) samples were not contaminated (Figure 1). Some broth enrichments had one isolate while others had more than one isolate.

The pie chart presented in Figure 2 shows the percentage distribution of 47 presumptive bacterial species from an initial 468 egg samples confirmed by MALDI-TOF analysis. The egg shell had the highest bacterial contamination at 43%, followed by the yolk at 30% and albumin at 27%. Within the scope of this study, only bacteria that cause gastrointestinal disorders were confirmed; the others were not further investigated.

In the dendrogram in Figure 3, the high discriminatory power of MALDI-TOF was used for the identification of different bacterial isolates with a high level of confidence. MALDI-TOF results are expressed as log (score) values ranging from 0 to 3 (0–100% pattern match). The higher the log (score) value, the higher the degree of similarity to a given organism in the reference database. A log (score) value of $\geq$2.00 can be considered an excellent probability for test organism identification at the species level; a value in the range 1.700–1.999 indicates probable genus identification and one in the range 0.000–1.6999 is not reliable for identification.[23,24] All 47 isolates had log (score) values of $\geq$2.00, which is the minimum threshold for secure species identification after MALDI-TOF spectral comparisons (Figure 3). None of the 47 isolates log (score) value was below 2.00. The protein profiles in normalised gel of confirmed bacteria showed the isolates' relationships with different egg brands, percentage relationship, grouping (from a–h) and the part of the egg from where they were isolated (Figure 3). The similarity pattern exhibited by the different bacterial species in each genus indicates that they are closely related with only slight variation in their protein profile that separates them into strains, as seen in their gel pattern. All the strains that have been grouped in the same cluster are assumed to belong to the same species (Figure 3).[26]

Three different serotypes of *Salmonella* were identified from 47 bacterial isolates that were previously identified by MALDI-TOF (Figure 3). *Salmonella* ser. Typhimurium represents 60% (28), *Salmonella* ser. Dublin 7% (3), and *Salmonella* ser. Braenderup 2% (1) of the identified isolates. The *Salmonella* species were present on the shells and in the albumin and yolk of eggs from different egg brands (Figure 3).

Discussion

The latest outbreak of *Salmonella* in Europe[27] emphasises the importance of this study. Consumers and health officials should be aware that hens' eggs may contain pathogenic bacteria. Isolating different bacteria in this study suggests that eggs sold by retailers in South Africa may be infected or contaminated with potential pathogenic bacteria; similar findings were previously reported by other researchers.[10] However, 27% of the eggs were uncontaminated, indicating that some retailers or producers practise hygienic measures to ensure the distribution of wholesome eggs to the public.

All the *Salmonella* isolates and representatives of the other bacteria that were subjected to MALDI-TOF analysis were present on the shell, in the albumin and in the yolk. This finding may suggest that these organisms may spread from contamination on the outside to the edible inside of the egg.[5]

Escherichia coli was isolated from samples AJ47 and AJ46 (marked as group 'f') and the relatedness in their protein profiles can be seen in Figure 3. This bacterium is a common enteric pathogen present in poultry[28,29] and it was isolated from only the shell surfaces of Brand 10 in this study. The shell is the most exposed part of the egg and *E. coli,* which is an enteric commensal, may have been transferred to the egg during the laying process or by trans-shell contamination with faecal material.[28,29]

Proteus mirabilis was isolated from the shell, albumin and yolk of egg in Brand 4 and the protein profiles in Figure 3 marked as 'c' of AJ35, AJ36 and AJ34 showed related bands. *P. mirabilis* contamination can be waterborne and this contamination may occur when the eggs are laid in a wet or damp environment because *P. mirabilis* is known to thrive in wet or damp environments and is also widely distributed in soil.[29-31]

Isolating *Klebshiella pneumoniae* from two different egg brands during this study is noteworthy. The protein profile of isolates from Samples AJ19 and AJ45 marked as 'b' in Figure 3 show that these bacteria are similar even though they are from two different egg brands. Intestinal infection caused by *Klebshiella* species can spread through the oral-faecal route. The presence of *Klebshiella* on eggs may be from more than one source because it easily spreads between hens kept in close contact in a hatchery via the respiratory system, faeces, fomites or even caretakers.[2,9] It is known that antibiotic resistance is very common in *Klebshiella* infections[32-34], making the present finding important to note so that appropriate measures can be taken to prevent disease occurrence.

Enterobacter cloacae is an organism present in the intestines of hens and may be transferred to the eggs during the egg laying process or by contamination of the egg shell with faeces.[29] If the eggs are kept in environments littered with contaminated faeces before and after packaging, this organism may penetrate through the shell to the albumin and yolk.[29] *E. cloacae* isolates from Samples AJ08, AJ09, AJ07, AJ21 and AJ20 grouped in group 'd' showing close grouping and relatedness although they were from two different egg brands (Brands 7 and 8).

Stenotrophomonas maltophilia is a pathogen of importance that can be found in water, soil and sewage or very humid conditions and may be spread to the chickens if they are kept in humid conditions.[29,35] *Stenotrophomonas maltophilia* that was isolated from the shell, albumin and yolk of an egg in Brand 7 marked as samples AJ12, AJ11 and AJ10 marked 'a' in Figure 3 had related bands in their protein profiles. The presence of this organism on the shell and inside the egg suggest that the organism may have been transmitted to the inside of the egg through faecal contamination of the shell.[29] Isolation of *S. maltophilia* from only one egg brand suggests that this infection may be limited to the farm from where these eggs were sourced.

The mass protein profile of the different *Salmonella* isolates was used to group them in three main clusters that look very similar with only a slight variation in the band pattern (Figure 3). All the strains that are grouped in the same cluster of the sequenced one are assumed to belong to the same species.[26]

Salmonella Typhimurium isolates were grouped into two separate clusters. The first cluster consists of *Salmonella* Typhimurium isolates from samples with code numbers AJ01–AJ06, AJ17–AJ18, AJ22–AJ31, AJ41, AJ43 and AJ44, which were grouped together as one cluster because of the similarity in their band patterns. The second cluster of *Salmonella* Typhimurium is made up of isolates with code numbers AJ13–AJ15, AJ31–AJ33 and AJ37 while another group that is made up of *S*. Dublin *and S*. Braenderup isolates with code numbers AJ38–AJ40 and AJ42, respectively, were all grouped together into the third separate cluster because of their similarity pattern (Figure 3). This result shows that all 32 *Salmonella* isolates were correctly identified at the genus to species level with a high level of confidence.

Egg Brands 7–11 were from a particular big chain supermarket that had several group stores with different brand names under which it retails its eggs. It is highly suggestive that eggs from this outlet may have been supplied by the same producer who packed the eggs under different brand names before distributing them to the smaller sales outlets as its marketing strategy.

Although the bacteria were widely distributed on the eggs, there was a higher percentage contamination on the egg shell, which is not surprising because the shell is the most exposed part of the egg which makes it most vulnerable to contamination. Contamination was next highest in the yolk and then in the albumin. The wide distribution of these bacteria on different parts of the egg may help to enhance their survival in the face of different environmental conditions.

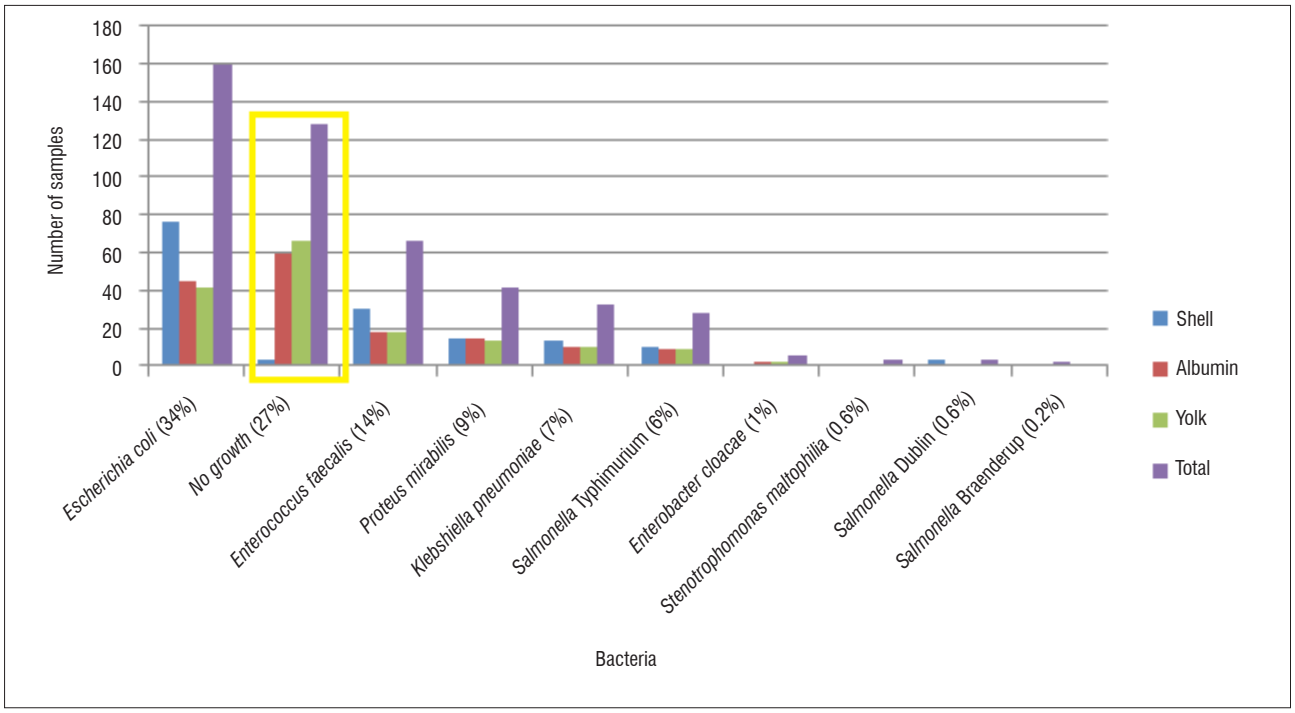


Figure 1: Distribution of different presumptive bacterial species isolated from the shell, albumin and yolk of retailed hens' eggs (n=468).

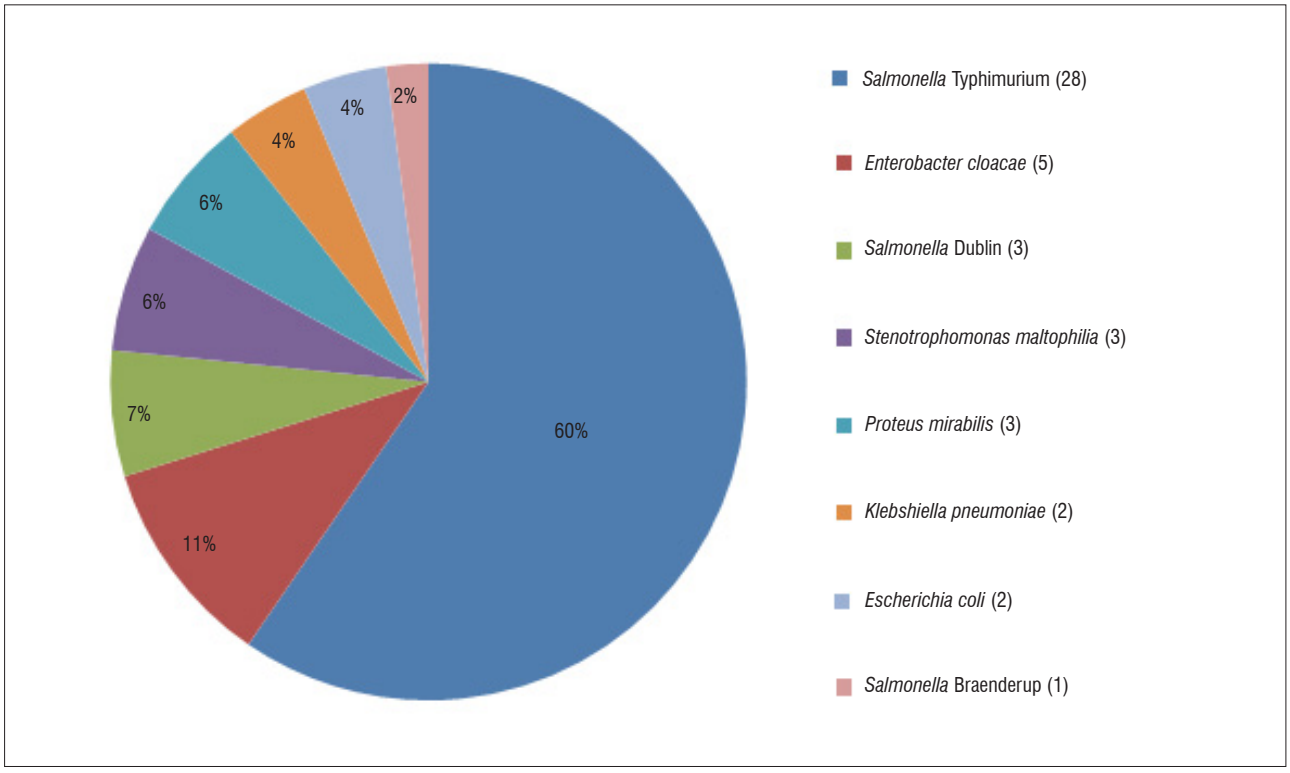


Figure 2: Distribution of presumptive bacterial species identified and confirmed by a Matrix-assisted Laser Desorption Ionisation Time of Flight (MALDI-TOF) analysis.

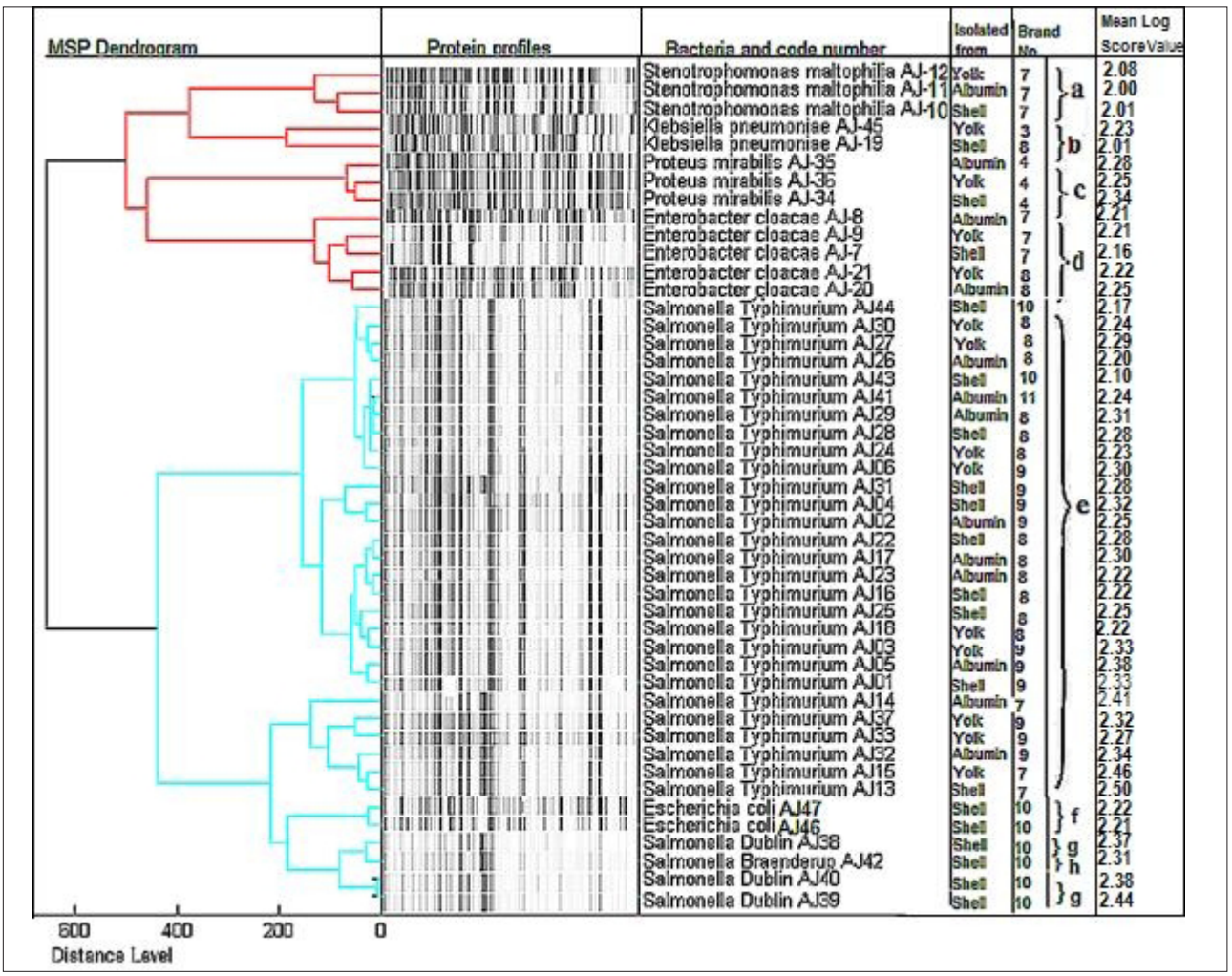


Figure 3: Dendrogram of mass spectral protein profiles in normalised gel of confirmed bacterial species showing log score values, isolates' relationships with different egg brands, percentage relationship grouping (from a–h) and the portion of the eggs from where they were isolated.

The isolation of different bacteria in retailed poultry eggs in this study may be an indication that some eggs that are sold to the public in Gauteng for consumption are not always of good quality. Some of these bacteria are members of coliforms, and coliform counts are used as indicators to evaluate the hygienic quality of raw foods.[2,3]

Researchers have previously isolated *Enterobacter, Klebshiella, Salmonella* and other bacterial species from egg shells in other places, that were similar with those isolated in this study.[9-11] This finding is important to note so that preventive measures can be put in place because of public health concerns.

Of a total of 13 egg brands analysed in this study, 5 egg brands were found to be contaminated by *Salmonella* species. *E. coli* was isolated from all the egg brands, which was not surprising because of its ubiquitous nature. The presence of *Salmonella* in eggs raises serious public health concerns and there may be a need to introduce hygienic regulations for producers and retailers regarding the hygienic quality of retailed eggs. Findings in this study also highlight the extent to which hens' eggs in South Africa are infected with different potentially pathogenic bacteria – information which has hitherto been very limited; our findings therefore address this knowledge gap.

Using contaminated unpasteurised eggs in different products poses serious health risks to consumers and could lead to multiple infections especially in immunocompromised persons such as those infected with the human immunodeficiency virus, further worsening the disease burden and contributing to an increase in mortality rates.[36,37] Therefore, it is advisable to use uncontaminated, preferably pasteurised, safe and wholesome eggs.

In South Africa, unlike in the USA, the EU and Canada[38-40], no law exists to regulate the content of raw eggs sold to consumers. Results from this study further emphasise the urgent need to introduce regulations on egg contents in South Africa.

The World Health Organization (WHO) recognises that control of *Salmonella* infection from poultry products can take place through public education, improvement of hygiene and control of infection in the birds themselves.[41] These measures can be applied to the other pathogens isolated in this study as a general control measure. Control can be achieved by observing strict biosecurity in the hatchery, breeding farm, environs, feed and water and during processing, which will ultimately protect the consumer.[41,42] WHO encourages the education of farmers and training of food handlers and consumers in food safety as the pivotal point of preventing salmonellosis. WHO and the Food and Agricultural Organization jointly encourage and enhance national, regional and provincial laboratories in the monitoring and surveillance of *Salmonella* transfer between food animals and humans, and in the coordination and response to outbreaks.[43]

A technical report was released by the European Food Safety Authority and the European Centre for Disease Prevention and Control on 27 October 2016 on a multi-country outbreak of *Salmonella* Enteritidis linked to unpasteurised eggs.[29] Confirmed and probable cases were reported in this outbreak; isolates of 112 of the confirmed and 148 of the probable cases belonged to two distinct genetic clusters. A fatal case linked to the outbreak was also reported. The eggs originated from a packing centre in Poland and were distributed to other countries in Europe. Restrictive measures to withdraw and stop orders for implicated eggs in the market were introduced while investigations to eliminate the source are ongoing.[27]

Conclusion

This study shows that some hens' eggs that are retailed in the Tshwane district of Gauteng Province, South Africa may contain potential pathogenic bacteria that may have public health consequences if the eggs are eaten undercooked, uncooked or used unpasteurised to prepare products containing raw hens' eggs. More efficient monitoring measures and even laws for public health concerns should be put in place in order to ensure that only uncontaminated, preferably pasteurised, safe and wholesome eggs are sold to consumers.

Acknowledgements

We acknowledge the staff of the Phytomedicine Programme, Department of Paraclinical Sciences, Faculty of Veterinary Science and the Department of Plant Pathology and Microbiology of the University of Pretoria for assisting with laboratory work and the National Veterinary Research Institute, Vom, Nigeria, for granting leave to A.R.J. to enrol for his PhD at the Faculty of Veterinary Science, University of Pretoria, South Africa. Funding of the project was through the research funds of F.S.B. and E.M.B.

Authors' contributions

A.R.J. was the main researcher, and was responsible for planning and executing the field experiments, the interpretation of the results and the write-up. E.M.B. co-supervised A.R.J., provided advice during the experimental work and the interpretation of results and edited the manuscript. F.S.B. supervised A.R.J., developed some of the methods executed by A.R.J., and assisted in the planning of the experimental work and the interpretation of results and edited the manuscript.

References

1. Holt J, Krieg N, Sneath PH, Staley J, Williams S. Facultative anaerobic Gram-negative rods. In: Holt J, editor. Bergey's manual of determinative bacteriology. 9th ed. Philadelphia, PA: Lippincott, Williams and Wilkins; 2000. p. 175–189.
2. Kornacki J, Johnson J. Enterobacteriaceae, coliforms and *Escherichia coli* as quality and safety indicators. In: Downes F, Ito K, editors. Compendium of methods for the microbiological examination of foods. 4th ed. Washington DC: American Public Health Association; 2001. p. 69–80. https://doi.org/10.2105/9780875531755ch08
3. Mercuri A, Cox N. Coliforms and Enterobacteriaceae isolates from selected foods. J Food Prot. 1979;42:712–714. https://doi.org/10.4315/0362-028X-42.9.712
4. Knape K, Chavez C, Burgess R, Coufal C, Carey J. Comparison of eggshell surface microbial populations for in-line and off-line commercial egg processing facilities. Poult Sci. 2002;8:695–698. https://doi.org/10.1093/ps/81.5.695
5. Cooper G, Nicholas R, Bracewell C. Serological and bacteriological investigations of chickens from flocks infected with *Salmonella enteritidis*. Vet Rec. 1989;125:567–572.
6. Gast R, Beard C. Production of *Salmonella enteritidis* contaminated eggs by experimentally infected hens. Avian Dis. 1990;34(2):438–446. https://doi.org/10.2307/1591433
7. Gast R, Holt P. Deposition of phage type 4 and 13a *Salmonella enteritidis* strains in the yolk and albumen of eggs laid by experimentally infected hens. Avian Dis. 2000;44:706–710. https://doi.org/10.2307/1593116
8. Humphrey T. Contamination of egg shell and contents with *Salmonella enteritidis*: A review. Int J Food Microbiol. 1994;21:31–40. https://doi.org/10.1016/0168-1605(94)90197-X
9. Davies R, Breslin M. Investigation of *Salmonella* contamination and disinfection in farm egg-packing plants. J Appl Microbiol. 2003;94:191–196. https://doi.org/10.1046/j.1365-2672.2003.01817.x
10. Musgrove M, Jones D, Northcutt J, Cox N, Harrison MA. Identification of Enterobacteriaceae from washed and unwashed commercial shell eggs. J Food Prot. 2004;67:1234–1237. https://doi.org/10.4315/0362-028X-67.11.2613
11. Musgrove MT, Northcutt JK, Jones DR, Cox NA, Harrison MA. Enterobacteriaceae and related organisms isolated from shell eggs collected during commercial processing. Poult Sci. 2008;87:1211–1218. https://doi.org/10.3382/ps.2007-00496
12. Jones D, Northcutt J, Musgrove M, Curtis P, Anderson K, Cox A. Survey of shell egg processing plant sanitation programs: Effects on egg contact surfaces. J Food Prot. 2003;66:1486–1489. https://doi.org/10.4315/0362-028X-66.8.1486
13. European Food Safety Association (EFSA). Preliminary report on the analysis of the baseline study on the prevalence of *Salmonella* in laying hen flocks of *Gallus gallus*. EFSA J. 2006;81:1–71. Available from: http://www.efsa.europa.eu/en/efsajournal/doc/81r.pdf
14. Food and Agricultural Organization (FAO) and World Health Organization (WHO). Risk assessments of *Salmonella* in eggs and broiler chickens: Interpretative summary. Microbiological Risk Assessment Series 1. Rome/Geneva: FAO/WHO; 2002.
15. Schroeter A, Ward L, Rowe B, Protz D, Hartung M, Helmuth R. *Salmonella enteritidis* phage types in Germany. Eur J Epidemiol. 1994;10:645–648. https://doi.org/10.1007/BF01719587
16. Henton M, Eagar H, Swan G, Van Vuuren M. Part VI: Antibiotic management and resistance in livestock production. S Afr Med J. 2011;101(8):583–586.
17. Van Vuuren M. Antibiotic resistance with special reference to poultry production. In: Proceedings of the 14th Conference of the OIE Regional Commission for Africa; 2001 January 23–26; Arusha, Tanzania. Paris: OIE Regional Commission for Africa; 2001. p. 135–146.
18. Hur J, Jawale C, Lee JH. Antimicrobial resistance of *Salmonella* isolated from food animals: A review. Food Res Int. 2012;45(2):819–830. https://doi.org/10.1016/j.foodres.2011.05.014
19. Pang T, Bhutta Z, Finlay B, Altwegg M. Typhoid fever and other salmonellosis: A continuing challenge. Trends Microbiol. 1995;3:253–255. https://doi.org/10.1016/S0966-842X(00)88937-4
20. Humphrey TJ, Whitehead A, Gawler AHL, Henley A, Rowe B. Numbers of *Salmonella enteritidis* in the contents of naturally contaminated hens' eggs. Epidemiol Infect. 1991;106:489–496. https://doi.org/10.1017/S0950268800067546
21. Kidanemariam A, Engelbrecht M, Picard J. Retrospective study on the incidence of *Salmonella* isolations in animals in South Africa, 1996–2006. J S Afr Vet Assoc. 2010;8:37–44. https://doi.org/10.4102/jsava.v81i1.94
22. Van der Zee H, Huis JH. Methods for the rapid detection of *Salmonella*. In: Wray C, Wray A, editors. *Salmonella* in domestic animals. Baltimore, MD: CABI Publishing; 2000. p. 300–390.
23. Bruker. MALDI Biotyper CA system clinical application for identification of microorganisms [document on the Internet]. c2014 [cited 2017 Mar 07]. Available from: http://www.aquilantscientific.com/assets/aquilantscientificie/products/brochures/442021/Bruker_MALDI-biotyper-brochure.pdf
24. Mellmann A, Cloud J, Maier T, Keckevoet U, Ramminger I, Iwen P, et al. Evaluation of matrix-assisted laser desorption ionization-time-of-flight mass spectrometry in comparison to 16s rRNAgene sequencing identification of non-fermenting bacteria. J Clin Microbiol. 2008;46:1945–1954. https://doi.org/10.1128/JCM.00157-08
25. Ward J. Hierarchical grouping to optimise an objective function. J Am Stat Assoc. 1963;58:236–244. https://doi.org/10.1080/01621459.1963.10500845
26. Ramirez-Castrillo'n M, Mendes SD, Inostroza-Ponta M, Valente P. (GTG)5 MSP-PCR fingerprinting as a technique for discrimination of wine associated yeasts? PLoS ONE. 2014;9(8), e105870, 8 pages. https://doi.org/10.1371/journal.pone.0105870
27. European Food Safety Authority (EFSA) and European Centre for Disease Prevention and Control (ECDC). Multi-country outbreak of *Salmonella enteritidis* phage type 8, MLVA type 2-9-7-3-2 and 2-9-6-3-2 infections. EFSA supporting publication. 2016;13(10), EN-1110, 20 pages. https://doi.org/10.2903/sp.efsa.2016.EN-1110
28. Nataro J, Kaper J. Diarrheagenic *Escherichia coli*. Clin Microbiol Rev. 1998;11(1):142–201.
29. Pattison M, McMullin P, Bradbury J, Alexander D. Poultry diseases. 6th ed. Toronto: Saunders Ltd Elsevier; 2008.

30. Schaffer J, Pearson M. *Proteus mirabilis* and urinary tract infections. Microbiol Spectr. 2015;3(5):10. https://doi.org/10.1128/microbiolspec.UTI-0017-2013

31. Nahar A, Siddiquee M, Nahar S, Anwar K, Ali S, Islam S. Multidrug resistant *Proteus mirabilis* isolated from chicken droppings in commercial poultry farms: Bio-security concern and emerging public health threat in Bangladesh. J Biol Heal Educ. 2014;2:2. https://doi.org/10.4172/2332-0893.1000120

32. Bodenstein J, Du Toit K. The susceptibility of *Staphylococcus aureus* and *Klebsiella pneumoniae* to naturally derived selected classes of flavonoids. In: Bobbarala V, editor. Antimicrobial agents. Rijeka: InTech; 2011. p. 73–84.

33. Fielding B, Mnabisa A, Gouws P, Morris T. Antimicrobial-resistant *Klebsiella* species isolated from free-range chicken samples in an informal settlement. Arch Med Sci. 2012;8(1):39–42. https://doi.org/10.5114/aoms.2012.27278

34. Sa MB, Ralph MT, Nascimento DCO, Ramos CS, Barbosa IMS, Sa FB, et al. Phytochemistry and preliminary assessment of the antibacterial activity of chloroform extract of *Amburana cearensis* (Allemao) A.C. Sm. against *Klebsiella pneumoniae* carbapenemase-producing strains. Evid Based Complement Alternat Med. 2014;2014, Art. #786586, 7 pages. https://dx.doi.org/10.1155/2014/786586

35. Alfieri N, Ramotar K, Armstrong P, Spornitz M, Ross G, Winnick J. Two consecutive outbreaks of *Stenotrophomonas maltophilia* (*Xanthomonas maltophilia*) in an intensive-care unit defined by restriction fragment-length polymorphism typing. Infect Cont Hosp Epidemiol. 1999;20(8):553–556. https://doi.org/10.1086/501668

36. Gordon NC, Wareham DW. Antimicrobial activity of the green tea polyphenol(-)-epigallocatechin-3-gallate (EGCG) against clinical isolates of *Stenotrophomonas maltophilia*. Int J Antimicrob Agents. 2010;36(2):129–131. https://doi.org/10.1016/j.ijantimicag.2010.03.025

37. World Health Organization (WHO). Global tuberculosis report 2015. 20th ed. Paris: WHO Press; 2015. Available from: http://www.who.int/tb/publications/global_report/gtbr15_main_text.pdf

38. Canada. Canada Consolidation Egg Regulations C.R.C. c. 284 [document on the Internet]. c2013 [cited 2013 Apr 26]. Available from: http://laws-lois.justice.gc.ca/PDF/C.R.C._c._284.pdf

39. European Union (EU). Commission regulation (EC) No 2073/2005 of 15 November 2005 on microbiological criteria for foodstuffs. Official Journal of the European Union. 2005:L338/21. Available from: http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32005R2073&from=EN

40. US Department of Health and Human Services, Food and Drug Administration (FDA). Prevention of *Salmonella enteritidis* in shell egg during production, storage and transportation; final rule. Federal Register. 2009;74(130):33030–33101. Available from: https://www.gpo.gov/fdsys/pkg/FR-2009-07-09/pdf/E9-16119.pdf

41. US Department of Health and Human Services, Food and Drug Administration (FDA). Prevention of *Salmonella enteritidis* in shell egg during production, 21CFR Parts 16 and 118. Federal Register. 2004;69(183):56823–56906. Available from: http://www.gpo.gov/fdsys/pkg/FR-2004-09-22/html/04-21219.htm

42. SAFE EGGS. Shell eggs pasteurization – the facts. Pluim Poultry Bull. 2008. Available from: http://www.safeeggs.com/safest-choice-pasteurized-eggs/how-we-pasteurize-eggs

43. World Health Organization (WHO). *Salmonella* (non-typhoidal): Fact sheet no. 139 [homepage on the Internet]. c2013 [updated 2017 Sep; cited 2017 Oct 10]. Available from: http://www.who.int/mediacentre/factsheets/fs139/en/

Bibliometric analysis of the development of nanoscience research in South Africa

AUTHORS:
Xolani Makhoba[1]
Anastassios Pouris[1]

AFFILIATION:
[1]Institute for Technological Innovation, University of Pretoria, Pretoria, South Africa

CORRESPONDENCE TO:
Xolani Makhoba

EMAIL:
xolani100@gmail.com

DATES:
Received: 07 Dec. 2016
Revised: 13 Mar. 2017
Accepted: 20 June 2017

KEYWORDS:
nanotechnology; R&D policy; bibliometric methods; collaboration

HOW TO CITE:
Makhoba X, Pouris A. Bibliometric analysis of the development of nanoscience research in South Africa. S Afr J Sci. 2017;113(11/12), Art. #2016-0381, 9 pages. http://dx.doi.org/10.17159/sajs.2017/20160381

ARTICLE INCLUDES:
× Supplementary material
× Data set

FUNDING:
None

Nanotechnology is a fast-growing scientific research area internationally and is classified as an important emerging research area. In response to this importance, South African researchers and institutions have also increased their efforts in this area. A bibliometric study of articles as indexed in the Web of Science considered the development in this field with respect to the growth in literature, collaboration profile and the research areas that are more within the country's context. We also looked at public institutions that are more active in this arena, including government policy considerations as guided by the National Nanotechnology Strategy launched in 2005. We found that the number of nanotechnology publications have shown a remarkable growth ever since the launch of the strategy. Articles on nanotechnology have been published in numerous journals, with *Electrochimica Acta* publishing the most, followed by *Journal of Nanoscience and Nanotechnology*. These publications fall within the traditional domains of chemistry and physics. In terms of the institutional profile and based on publication outputs over the period reviewed, the Council for Scientific and Industrial Research is a leading producer of publications in nanotechnology, followed by the University of the Witwatersrand – institutions that are both based in the Gauteng Province. There is a high level of international collaboration with different countries within this field – the most productive collaboration is with India, followed by the USA and China, as measured through co-authorship.

Significance:

- Nanotechnology as a field of research is experiencing rapid growth and there is a need to understand progress from a South African perspective.

Introduction

Innovations linked to science and technology developments are believed to be promising contributors to economic progress internationally. In response to this promise, South Africa adopted the National Research and Development Strategy[1] in 2002 to enable the transition from a resource-based economy to a knowledge economy, which was followed by the Ten-Year Innovation Plan towards a knowledge-based economy[2]. This strategy was viewed as pivotal to the achievement of national goals, including enhanced economic growth, industrial competitiveness, as well as social and developmental aspirations.[3] To this end, a number of targeted interventions and investments were made in specific fields of science, one of which was nanotechnology. Nanoscience and nanotechnology (N&N) is now widely recognised as an area of science and technology that promises to bring many scientific breakthroughs in the coming years, which will have a meaningful impact on the economy. According to Lux Research[4], the revenue from nano-enabled products worldwide as of 2014 stood at USD1.6 trillion – a growth of 90% in 2 years from USD850 billion in 2012. The application of nanomaterials occurs in many industries such as cosmetics, construction and electronics. It is important to note, as Reiss and Thielmann[5] have pointed out, that nanotechnology does not constitute a product specifically, but is in most cases integrated in a large variety of different applications in a large number of industrial sectors. Nanotechnology can therefore be understood as an enabler of innovative technologies and applications by substituting and improving existing products or leading to fundamentally new products.[5] As a result, various governments have been investing in the development of nanotechnology in their respective countries.

The Organisation for Economic Co-operation and Development (OECD) defines nanotechnology as a set of technologies that enable the manipulation, study or exploitation of structures and systems of typically less than 100 nanometres in size. It is believed that developments in N&N have the potential to affect virtually every area of economic activity and aspect of daily life.[6] This wide-ranging effect is because of the technology's ability to contribute to the development of novel materials, devices and products. The number of products, and the diversity of nanomaterials and nanosystems, is predicted to increase rapidly in the coming decade as a result of continuous innovation in many sectors. Nanotechnology can be applied in many commercial products in areas such as health (especially drug-delivery), energy, food packaging and water purification systems. It is evident that no work has been done to evaluate the development in this field since the N&N strategy was introduced in South Africa. The aim of this study was to establish publication and citation trends in the area of nanoscience. Secondly, we aimed to assess collaboration and the most productive countries and institutions publishing in the area of nanoscience; and, finally, we aimed to determine the intellectual structure of the nanotechnology research in terms of subject area. South Africa's investment in N&N research, and in other areas, is seen as a means to moving towards the creation of a knowledge-based economy, and as such, the country's progress in this regard. In terms of terminology, nanoscience and nanotechnology are used interchangeably throughout the article, as is the practice in this research community.

Background

Nanotechnology in South Africa and other developing countries

Based on the potential economic benefits and developments internationally, the South African government made some policy interventions to stimulate the development of nanotechnology, after the establishment of the National Nanotechnology Strategy[7] together with the stakeholder community under an organisation called the South African Nanotechnology Initiative (SANi). The goals of the strategy were to 'support long-term research that will lead to the fundamental understanding of nanomaterials' and a more ambitious target of supporting the 'creation of new and novel devices for application in various areas such as health, water and energy'. To support these objectives, the government committed to several investments, including inter alia:

- Establishing Nanotechnology Innovation Centres with a mandate to build capacity to develop commercial nano-enabled products. Two of these centres have been established: one is based at Mintek and the other at the Council for Scientific and Industrial Research (CSIR), both of which are situated in the Gauteng Province.
- Providing ring-fenced grants to researchers through the National Research Foundation for purchase of nanotechnology-related research equipment under the National Equipment Programme funding instrument.
- Initiating and funding the establishment of a taught master's degree programme in N&N which is currently offered by four universities: the University of the Western Cape, Nelson Mandela University, the University of Johannesburg and the University of the Free State.

These initiatives and other investments need to be contextualised as South Africa entered the N&N field later than developed economies. Government support through policy intervention and funding in this field is acknowledged.[8] Although the investments were well considered, they may not be sufficient to allow South Africa to play a leadership role in this research area. Research done by Pouris[9] indicated that the number of articles in the field produced in South Africa had increased from just 12 in 2000 to 57 by 2005, with the University of the Witwatersrand being a leading producer. Pouris[9] further highlighted the absence of science councils in the top producing category at the time, with N&N research being largely driven by individual academics rather than by a coordinated national approach. Pouris[9] also showed that the USA was a top collaborating country with South Africa. It is further noted that the number of N&N core journals currently indexed by the Science Citation Index is now at 83 – a number that has been increasing steadily since the area started gaining recognition in the 1990s.

Looking broadly at the participation of African researchers in nanotechnology, it has been noted that participation is very low and fragmented. Generally, the growth in nanotechnology in the least developed countries, including those in Africa, has been very slow; while individuals have shown interest in this field, there is no practical plan for the advancement of this field and as a result nanotechnology remains an area of academic research.[10] Ezema et al.[10] further point out that the BRICS nations produced a substantial number of publications, but that outside this grouping, there was very little activity in developing countries, especially in Africa, partly because of a lack of research infrastructure and facilities to carry out N&N research and thus these nations are classified as nanotechnology dormant. Maclurcan[11] noted that there are niche areas within N&N that are of benefit to developing countries such as India and South Africa, which – because of their high prevalence of tuberculosis (TB) infections – have programmes for developing nanotechnology-enabled TB diagnostic kits and improved TB drug-delivery systems. These specific circumstances provide a clear niche area for these countries to establish competitive expertise.

Bibliometrics

Bibliometric approaches are increasingly used for the assessment of scientific disciplines and provide useful information for those who fund research. It has been established that increased investment should lead to an increase in the number of publications; this investment also is more positive if it is made from the public sector rather than from the private sector.[12] Bibliometrics is defined as the application of mathematics and statistical methods to communication media. Bibliometrics have been used extensively to evaluate research progress quantitatively, for example in studying research output for specific countries[13-15] in a continent or a region such as Africa or southern Africa[16,17] or sometimes for a research discipline such as nanotechnology[9,18] or even a single institution[19]. One problem identified in bibliometric research is the classification and delineation of fields, for example, for new and emerging areas like biotechnology and nanotechnology the definitions differ among different application areas, including which areas it encompasses. Leydesdorff[20] further states that delineation using core journals as done by Clarivate Analytics' (formerly Thomson Reuters) Web of Knowledge classification may not account for research published in multidisciplinary journals. In areas such as these, delineation is achieved by using keywords to extract the relevant publications from the database, as in this case; another approach is content analysis or using only core journals or a combination thereof. In a multidisciplinary area like N&N in which literature is scattered, delineation is more important and only keywords are able to extract the relevant publications as there are relatively few established core journals and the literature tends to be published in traditional journals such as those in the fields of chemistry and physics.

In this study, bibliometrics analysis was performed on the data obtained from a citation index database. A citation index is defined as an ordered list of cited articles, each of which is accompanied by a list of citing articles.[21] There are many academic citation indexes used in bibliometrics research such as CiteSeer, Google Scholar and Elsevier's Scopus depending on the context of the study, but the most commonly used is Clarivate Analytics' (formerly Thomson Reuters) Web of Science™ (WoS).

Methodology

Data were extracted from articles in nanoscience and nanotechnology published over an 11-year period starting in 2005. The year 2005 was chosen because it coincides with the launch of the N&N strategy in South Africa. It would not have added value to go back further than 2005 as the study mentioned earlier[9] covers the period up to 2005. The Science Citation Index offered by WoS was used for the search of journal publications.[22] A major strength of this database is that it provides sufficient coverage of the most important and influential journals and core literature internationally. For the calculation of activity index and percentage share, the total number of publications was obtained using InCites™ – an analytical tool provided by Clarivate Analytics which uses the same underlying WoS data.

A keyword-based search is the preferred route for a relatively new and multidisciplinary field like N&N, as there are a high number of relevant publications that can be hidden in multidisciplinary journals. Based on this preference, the WoS core collection database was used, which allows for a keyword-based search. There are many possible search strategies, from the simple use of the nano* wildcard to the more elaborate approach of a modular search strategy as described by Porter et al.[23] A simplified methodology by Maghreb et al.[24] that recognises that not all words that start with 'nano' refer to nanomaterials and that there are some nanomaterials that do not have keywords containing the nano- prefix (such as quantum dots and fullerenes) was used. This methodology was used because of its simplicity and very accurate selection of nanotechnology articles. The following keywords were used: nano* NOT nano2 NOT nano3 NOT nanog* NOT nanosecond* NOT nanomol* NOT nanogram* NOT nanoplankton* OR "atom* scale" OR "atomic layer deposition*" OR "giant magnetoresist*" OR graphen* OR dendrimer* OR fulleren* OR "c-60" OR "langmuir blodgett*" OR mesopor* OR "molecul* assembl*" OR "molecul* wire*" OR "porous silicon*" OR "quantum dot*" OR "quantum well*" OR "quantum comput*" OR "quantum wire*" OR qubit* OR "self assembl*" or supramolecul* OR supermolecul* OR "ultrathin film*" OR "ultra thin film*". In this case, a top down keyword search and Boolean operators were used. This approach enabled extraction of all articles containing

keywords known to be used in nanotechnology publication. The operators OR and NOT were used to exclude those articles that may include the nano- prefix but are not related to nanotechnology. Only research articles were included; other publication types such as book chapters and proceedings were not included.

In this article, a descriptive approach is used to identify the trends based on the publication data. The analysis focuses on countries with which South Africa collaborates the most on research, the most prolific research institutions in N&N research, the top journals selected by South African researchers as well as the subject categories, as N&N is interdisciplinary.

Results and discussion

The aim of this study was to investigate the productivity and intellectual structure of N&N in South Africa since the launch of the country's N&N strategy. The results of the bibliometric study obtained from the analysis of the 2928 records extracted from the WoS indexing database and published in the 11-year period between 2005 and 2015 are presented.

The first part considers the publication and citation trend; the second part, the collaboration and the most productive countries and institutions; and, finally, the intellectual structure of the nanotechnology research is considered. A search of patents using the same keywords as used for the publication search revealed that South Africa has been granted fewer than 10 nanotechnology patents by the USPTO over this period; as a result of this small number, patents did not form part of the study. The number of patents was confirmed using the OECD database[25] and, based on the available data, further analysis was abandoned as it would not have added further insight to this study.

The nanotechnology publication trend in South Africa

Figure 1 shows the publication trend of nanotechnology articles in South Africa between 2005 and 2015. Figure 1 shows an increasing trend from 2005 when only 62 publications were produced to a total of 597 in 2015 – an average annual growth rate of 22%. The number of publications in N&N have grown substantially, but to put this growth in context, the total publication number in South Africa grew from 6408 in 2005 to 15 468 in 2015 – equating to an average annual growth rate of 14%. Therefore, N&N publications grew at a faster rate than the growth of all publications in South Africa, which equates to an average annual growth rate of 0.36%. Regardless, this growth in N&N outputs is almost tenfold.

To ascertain the significance of the growth in nanotechnology articles, the percentage of nanotechnology articles to overall articles from the country during the period was calculated. It is clear from Figure 1 that this percentage grew from less than 1.5% in 2005 to just above 4.5% in 2015 – a threefold increase and much higher than the overall growth of articles generally. This growth is also distinct from that of energy publications, for example, which, according to a recent study, has remained stagnant since 2008.[26]

The citation trend of nanotechnology publications in South Africa

Figure 2 shows the citation trend of nanotechnology articles in South Africa from 2005 to 2015. An observation of the citations of the articles reveals an even more impressive picture than that of publication growth (Figure 2). The citations have grown from 18 citations in 2005 to 7229 in 2015 – an average annual growth rate of 72%, which is quite impressive considering that the overall growth in citations in the country was quite low. Citations for South African publications overall have not grown much, with a growth of 1.81% between 2005 and 2010 compared with 251% for N&N over the same period. The nanotechnology articles generated by South African researchers have relatively high visibility.

To determine the significance of the growth in citations in nanotechnology articles, we calculated the share of citations to N&N articles as a percentage of citations to all articles produced in South Africa. As seen from the secondary axis in Figure 2, this share grew from just 0.04% in 2005 to above 1.6% in 2015. The growth in citations to nanotechnology articles was far higher than the growth of total citations to all South African publications over the same time.

Table 1 shows the publication and citation trends of nanotechnology articles in South Africa from 2005 to 2015. Citations represent the impact and the influence of articles, so this high citation trend indicates that the nanotechnology articles generated by South African researchers have a high impact.

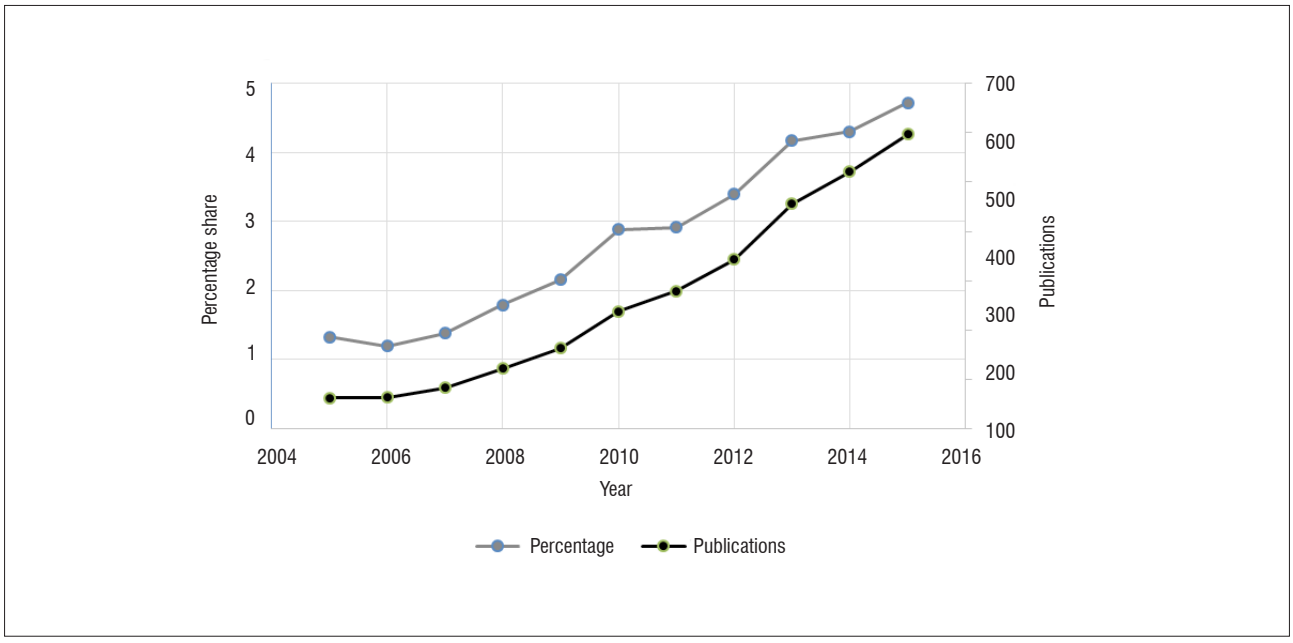


Figure 1: The publication trend as well as percentage of nanotechnology articles relative to the total number of articles in South Africa from 2005 to 2015.

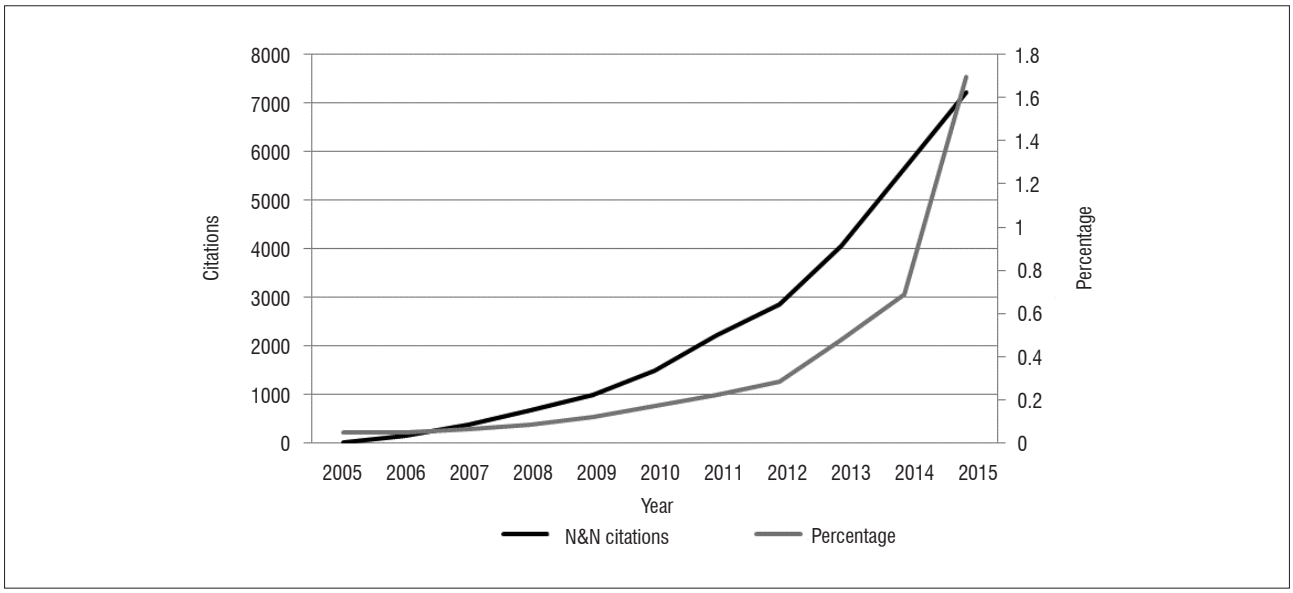


Figure 2: The citation trend as well as percentage of citation of nanoscience and nanotechnology (N&N) articles relative to the total number of citations in South Africa from 2005 to 2015.

Table 1: South Africa's publication and citation trends in the field of nanoscience and nanotechnology (N&N) and growth rates for 2005–2015

Year	Total number of articles	Number of N&N articles	Growth rate (%)	Total citations	Citations for N&N articles	Growth rate (%)
2005	6409	62		129 563	18	
2006	7310	64	3.22	131 899	156	940
2007	8569	83	29.68	129 236	365	133.97
2008	9560	122	46.98	143 538	671	83.84
2009	10 602	163	33.60	137 147	997	48.84
2010	10 936	238	44.17	139 804	1477	48.14
2011	12 563	279	17.22	126 577	2211	49.69
2012	14 311	343	22.94	121 541	2858	27.78
2013	14 890	456	32.94	94 893	4048	41.63
2014	16 260	521	14.25	75 991	5662	39.87
2015	17 246	597	14.58	35 174	7229	27.68

South African output in N&N during the period 2005–2015 included 2928 publications. With respect to publication output, the nanotechnology area grew at an average annual growth rate of 25.95% and the citation growth was much higher at an average annual growth rate of 144%. This growth is in line with the international growth in N&N which has been reported[27] to be at 23% and quite significantly higher than the average annual growth rate of publications in South Africa in general, which is 14%. The growth rate of N&N articles was 3.2% in 2006, and accelerated sharply to almost 30% in 2007, and up to 46% in 2008. This area of research has been growing at a fast pace and finally stabilised at about 14% in 2014/2015. It will be interesting to observe how the growth evolves in the next few years or whether 14% is the 'new normal', as it is in line with the growth of South African publications in general. The citations to N&N articles grew from a small base of only 18 citations in 2005 to 156 citations in 2006 – a 940% increase. This growth decelerated and finally settled at about 48% in 2009–2011. Citations saw a growth of 27% in 2015 – not as high as in earlier years. The growth of nanotechnology – both in terms of articles published and the citations they have received – is phenomenal and indicates that researchers are spending resources in this field of research and that the government support initiatives are successfully stimulating interest in this area.

Table 2 shows the publication trend of nanotechnology articles in other selected countries for comparison purposes from 2005 to 2015. The ratio of N&N articles produced to total articles in all areas is a useful means of comparison, as is the activity index. The totals were obtained using InCites™.

Table 2: Number of publications in the field of nanoscience and nanotechnology (N&N) compared with those of selected countries, 2005–2015

	N&N articles	Total articles	Ratio	Activity index
World	996 083	15 914 248	0.062	1
China	220 413	2 335 407	0.094	1.50
USA	165 691	6 205 056	0.027	0.43
India	49 981	603 489	0.083	1.32
Russia	23 755	372 700	0.064	1.01
Brazil	11 878	439 444	0.027	0.43
Egypt	5236	80 346	0.065	1.04
South Africa	2928	122 126	0.024	0.38

Table 2 shows that South Africa, compared with the other BRICS countries, Egypt and the USA, had the least number of N&N articles, with only 2.4% of publications being in the nanoscience field. South Africa's world share of N&N articles over this period was 0.29% – far below South Africa's world share of total articles of 0.79%. This finding may indicate that N&N research output has most likely not reached its full potential in South Africa. N&N is a high growth research area internationally, and the growth in the number of South African publications is not keeping up with the growth in the overall growth of the field worldwide. Egypt produced 5236 publications over the same period – placing this country in a leading position in N&N on the African continent. Chen et al.[28] found that Russia, India and China showed rapid growth in N&N publications from 2000 to 2007 of about 12.8 times in China, 8 times in India and 1.6 times in Russia. What is apparent is that N&N is growing very rapidly while the growth in South Africa has been recorded to be below that of other countries. Appelbaum et al.[29] confirm the importance of nanotechnology in China with a number of state-led interventions; the Chinese government is also investing an estimated USD200 million per year in this field, making it second only to the USA in terms of investment. The USA launched its Nanotechnology Initiative in 2000, with an annual budget for 2016 estimated at just over USD1.4 billion, while the Chinese government followed a year later with its own initiative in 2001; in both of these countries, the N&N initiatives were accompanied by big dedicated budgets.[30] South Africa only launched a similar initiative in 2005 with what can be described as a very modest budget. It is interesting that China, India, Russia and Egypt show figures above 1, which is an indication that these countries have placed more emphasis in N&N. Despite a higher budget, the USA – along with South Africa and Brazil – have activity indices of less than 1, showing no specialisation in this field. Besides looking at the number of publications, another indicator for the comparison with other countries is the activity index, which was first described by Frame[31]. The index indicates the country's share in world publications in a particular field of science relative to the overall share in world total publications. This index has been used quite recently in other studies such as the one by Makhoba and Pouris[32] in which different scientific priority areas in South Africa were compared. The activity index is zero when the country holds no publications in that discipline and it is equal to 1 if the country's share in the discipline equals the country's share in all fields indicating no specialisation and above 1 when a positive specialisation is observed.

Collaboration and publication profiling of South Africa

Collaboration with researchers from other countries was examined for the nanotechnology field as indicated by co-authorship of research papers. The top 20 countries based on the co-authored articles are given in Figure 3.

Figure 3 indicates that collaboration in the nanoscience and nanotechnology research area is the highest with India, followed by the USA and China. In terms of numbers, South Africa collaborated on publications during the 11 years with India (266 publications), the USA (190 joint publications), China (127 publications), Germany (123 publications) and England (112 publications). The aggregate share of collaboration with these top five collaborating countries (818 publications) is actually 28% of the total number of all publications. The USA is a global leader in the nanotechnology field, and consequently is a key country with which to collaborate. The presence of India as the top collaborating partner may come as a surprise to some, but South Africa has strong historical ties with India which pre-dates their involvement in the BRICS grouping. In a

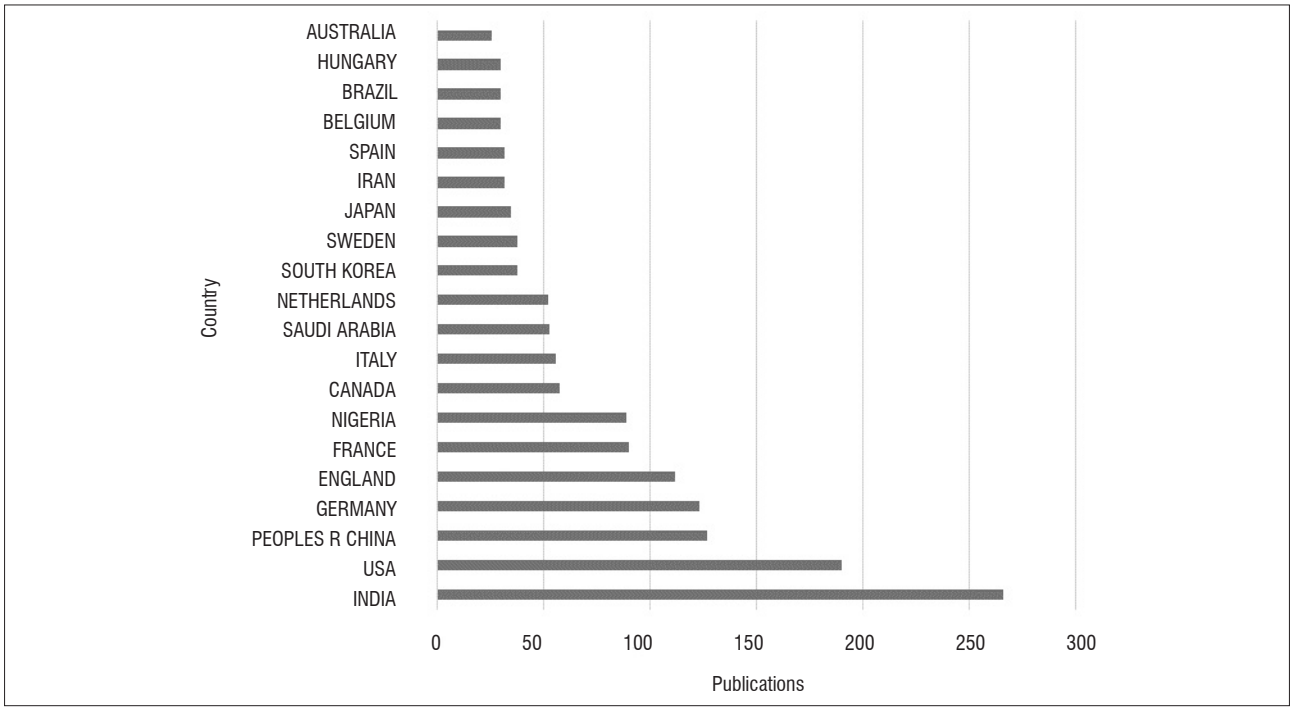


Figure 3: Nanotechnology collaboration profile of South Africa with other countries.

study of the BRICS countries, Finardi[33] found that South Africa and India have the strongest collaborations compared with other countries in the BRICS grouping. This high collaboration was attributed to both countries belonging to the Commonwealth, both once being part of the British Empire and both having English as one of their official languages.

Collaboration results are not expected to add up to 100% because most of the articles are authored exclusively by South Africans with no external collaborator. It is important to note that South Africa has fewer collaborations within the continent, with only nine collaborating countries in Africa. Collaboration with other African countries produced a total of 130 publications – accounting for only 5.57% in this field in South Africa. Collaboration in this area with Nigeria – which is South Africa's biggest collaborator in Africa – produced only 58 articles which is low compared with India, for example. This result is hardly surprising as it has been established in past studies that African researchers prefer to collaborate with researchers outside of the continent. Pouris and Ho[17] found that South African researchers have been increasingly publishing with international partners, with a growth of 66% in the 5 years between 2007 and 2011, making up a total of 54% of all articles produced in the country. This growth is on the back of the findings by Boshoff[34] that collaboration with neighbouring countries in the Southern African Development Community remained stagnant at 3% for the 3 years between 2005 and 2008.

Table 3: The top 24 nanotechnology journals used by South African researchers

Journal	Impact factor	Record count	Country
Electrochimica Acta	4.803	62	England
Journal of Nanoscience and Nanotechnology	1.338	61	USA
International Journal of Electrochemical Science	1.692	54	Serbia
Materials Science	0.143	51	Ukraine
Journal of Applied Polymer Science	1.866	47	USA
RSC Advances	3.289	41	England
Journal of Alloys and Compounds	3.014	39	Switzerland
Applied Surface Science	3.150	36	The Netherlands
Physica B: Condensed Matter	1.352	33	The Netherlands
Journal of Materials Science	2.302	32	USA
Polyhedron	2.108	32	England
International Journal of Hydrogen Energy	3.205	30	England
South African Journal of Science	0.902	29	South Africa
Polymer	3.586	26	England
Carbohydrate Polymers	4.219	25	England
Journal of Power Sources	6.333	25	The Netherlands
Materials Chemistry and Physics	2.101	25	Switzerland
Electroanalysis	2.471	24	Germany
Optical Materials	2.183	23	The Netherlands
Journal of Luminescence	2.693	23	The Netherlands
Journal of Nanoparticle Research	2.101	22	The Netherlands
Journal of Photochemistry and Photobiology A: Chemistry	2.477	21	Switzerland
New Journal of Chemistry	3.277	21	England
Thin Solid Films	1.761	21	The Netherlands

Nanotechnology journals used by South African researchers

Table 3 highlights the top 24 journals in which N&N publications by South African researchers between 2005 and 2015 are to be found. Included is the journal impact factor as published in the Journal Citation Reports of Clarivate Analytics (formerly Thomson Reuters) for the year 2015. South African authors published most in the journal *Electrochimica Acta* (62 articles) followed by the *Journal of Nanoscience and Nanotechnology* (61 articles) and lastly the *International Journal of Electrochemical Science* (54 articles). The top 24 journals have a significant 27% (803) share of the total publications, indicating a wide scatter of literature in nanotechnology with no clear concentration of articles in one group of journals. In this list, the low representation of South African journals is observed, indicating a preference for international journals, with only the *South African Journal of Science* represented in the top 24. The most used language in the top 24 journals is English, with the rest of the journals being bilingual.

The top 24 journals are mostly from England and the Netherlands, with each of these countries having seven journals on the top 24 list. The USA – with which South African researchers collaborate the most – has only three journals on the top 24 journals list. Therefore, it is clear that South African researchers are leveraging international resources to catch up in nanotechnology research. Additionally, they attempt to increase the visibility of their research by publishing most of their articles in journals from European countries whilst still collaborating with the USA. In terms of impact factor, the journals range from *Materials Science* from the Ukraine with the lowest impact factor of 0.143 to the *Journal of Power Sources* from the Netherlands with the highest impact factor of 6.333. The average impact factor for the journals listed in Table 3 is 2.60, as published in the 2015 Journal Citation Reports.

Subject area structure

An analysis of N&N publications and their subject categories was done to understand the subject area structure (Table 4).

The highest number of publications in N&N by South African researchers fall into the three traditional domains: Chemistry, Physics and Materials Science. The top three subject areas account for 88.9% of the country's publications in the field of nanoscience and nanotechnology. It is clear therefore that most of the nanotechnology applications are in Physics, Chemistry and Materials Science. The priority application areas that are emphasised in the South African N&N strategy – such as water, health and energy – seem well presented, although not to the extent envisaged in the strategy. It is also interesting that there is a broad spread in terms of the application of N&N including in areas such as Mechanics, Optics and Geology. The results are not expected to add up to 100% as the articles can belong to more than one research domain and only the top 26 fields are reported.

Publishing organisations

Analysis of the most prolific institutions indicates a dominant contribution of institutions based in the Gauteng Province, with three of the top five institutions based in this province (Table 5). It is unclear whether this was a deliberate policy objective or a coincidence; the N&N strategy is silent on this issue. The CSIR accounted for 16.4% of publications in the N&N area, followed by the University of the Witwatersrand at 14.2%. Government, through the Department of Science and Technology, established and funded the Nanotechnology Innovation Centre, which is sometimes referred to as the National Centre for Nano-Structured Materials based at the CSIR, which explains the CSIR's leadership role in this field. The centre at the CSIR called the National Centre for Nano-Structured Materials is one of two nanotechnology centres established by the government in 2007. Its focus is on the development of new nanotechnology-enabled materials, with applications in the manufacturing, water and health sectors. The National Centre for Nano-Structured Materials as a government-funded institute makes available high-tech instrumentation to other researchers in South Africa. The other national centre is based at Mintek and focuses on the fields of sensors, biolabels and water nanotechnology. Interestingly, the National Research Foundation also features strongly in the third position with most of the output attributed to the national facilities, notably iThemba Labs.

Table 4: Nanotechnology subject area map in South Africa

Research area	Record count	Percentage of total
Chemistry	1089	37.19%
Materials Science	910	31.08%
Physics	605	20.66%
Science Technology and Other Topics	383	13.08%
Polymer Science	338	11.54%
Engineering	289	9.87%
Electrochemistry	278	9.50%
Pharmacology Pharmacy	105	3.59%
Environmental Sciences Ecology	100	3.42%
Energy Fuels	91	3.11%
Metallurgy and Metallurgical Engineering	82	2.80%
Optics	76	2.60%
Biochemistry Molecular Biology	75	2.56%
Crystallography	57	1.95%
Water Resources	57	1.95%
Instruments and Instrumentation	56	1.91%
Biotechnology and Applied Microbiology	54	1.84%
Mechanics	48	1.64%
Thermodynamics	44	1.50%
Nuclear Science and Technology	31	1.06%
Spectroscopy	28	0.96%
Mathematics	26	0.89%
Biophysics	22	0.75%
Toxicology	22	0.75%
Microbiology	21	0.72%
Geology	20	0.68%

In terms of inter-institutional collaboration, the most prolific institutions in N&N research in South Africa collaborate with each other extensively, as indicated by the number of co-authored articles. The CSIR, with a total of 479 publications, collaborated with the University of Johannesburg (80), University of Pretoria (69) and Tshwane University of Technology (56) nationally, and the King Abdulaziz University in Saudi Arabia (21) and the University of Malawi (11), internationally. The first three institutions – like CSIR – are in Gauteng, with close proximity facilitating collaboration. The CSIR is also a national facility which provides access to analytical equipment to researchers countrywide, and therefore the high number of papers co-authored with researchers from other South African institutes is not surprising. The University of the Witwatersrand, with a total of 420 publications, collaborated with the National Research Foundation (169), University of Johannesburg (54) and CSIR (46) nationally, and, internationally, with the Federal University of Paraná in Brazil (18), Tamkang University in Taiwan (11), University of Ulster in Northern Ireland (11) and the University of Malawi in Malawi (9). Clearly an argument can be advanced that institutions from Gauteng – South Africa's economic hub – are leading N&N research and it appears that there is a close level of collaboration between them. It would seem that, despite its relatively small scientific size, Malawi (through the University of Malawi) has a very close relationship with the two top South African institutes in N&N research.

Table 5: Most prolific organisations in nanotechnology publishing in South Africa

Organisation	Record count	Percentage of total
Council for Scientific and Industrial Research	479	16.36%
University of the Witwatersrand	416	14.21%
National Research Foundation	397	13.55%
University of Johannesburg	381	13.01%
University of the Western Cape	269	9.19%
Rhodes University	264	9.02%
University of KwaZulu-Natal	249	8.50%
University of the Free State	244	8.33%
University of Pretoria	218	7.45%
Stellenbosch University	215	7.34%
Tshwane University of Technology	158	5.40%
University of South Africa	133	4.54%
University of Cape Town	120	4.10%
North West University	104	3.55%
University of Zululand	87	2.97%
Nelson Mandela University	75	2.56%
Mintek	71	2.43%
Cape Peninsula University of Technology	50	1.71%
Vaal University of Technology	43	1.47%
Durban University of Technology	42	1.43%
University of Fort Hare	28	0.96%
Walter Sisulu University of Technology	24	0.82%
University of Limpopo	23	0.79%
Sasol Technology	15	0.51%

Tables 5 and 6 show an obvious lack of representation from the private sector, with only Sasol Technology publishing in this field. This finding does not augur well for the future commercialisation of nanotechnology-enabled products in the country as envisaged in the N&N strategy. Perhaps this result could be explained in that the private sector may prefer the patenting route or retaining trade secrets rather than disseminating their research findings through publications.

Conclusion

As a result of the growing technological significance and expected economic contribution, N&N in South Africa has been thoroughly analysed through bibliometric methods. The research output in this field is showing a steady increase in South Africa since the introduction of the National Nanotechnology Strategy in 2005 and its associated government support. However, to put this increase into context, there is a need to look at some other countries during the same period, such as Egypt which produced almost double the number of publications to that of South Africa. Looking at the other BRICS countries, China produced 220 413, India 49 981, Russia 23 755 and Brazil 11 878 articles over this period. The USA produced a total of 165 691 articles during this period, in line with their big budget for N&N which was projected to reach USD1.5 billion in 2016.

Table 6: The two most prolific institutional publishers of nanoscience research in South Africa and their collaborative partners

Council for Scientific and Industrial Research (South Africa)		University of the Witwatersrand (South Africa)	
Collaborating institution	**Number of publications**	**Collaborating institution**	**Number of publications**
University of Johannesburg (South Africa)	80	National Research Foundation (South Africa)	169
University of Pretoria (South Africa)	69	University of Johannesburg (South Africa)	54
Tshwane University of Technology (South Africa)	56	Council for Scientific and Industrial Research (South Africa)	46
University of the Western Cape (South Africa)	52	Mintek (South Africa)	27
University of the Witwatersrand (South Africa)	46	University of KwaZulu-Natal (South Africa)	18
University of the Free State (South Africa)	45	Vaal University of Technology (South Africa)	18
University of South Africa (South Africa)	31	Federal University of Paraná (Brazil)	13
National Research Foundation (South Africa)	50	Tamkang University (Taiwan)	11
King Abdulaziz University (Saudi Arabia)	21	University of Ulster (Northern Ireland)	11
University of KwaZulu-Natal (South Africa)	19	University of Malawi (Malawi)	9
North West University (South Africa)	17	University of South Africa (South Africa)	9

This discrepancy may be partly because of South Africa's low budget for R&D in general but also lack of prioritisation of N&N in the country. This lack of prioritisation is evident when one looks at the low activity index, with Egypt taking a leading position in this field on the African continent. So, in conclusion, the policy has institutionalised N&N research within science councils and universities, as opposed to what was the case before the introduction of the strategy. However, South Africa has the potential to produce more output in the N&N field, based on the comparison of output from other science areas. In a comparison of OECD countries, it was found that in 2014 the gross expenditure of R&D as a percentage of GDP was 0.73% for South Africa, whereas the average for OECD countries is 2.38%; interestingly, for China, this value stood at 1.93% for 2014.[6] Therefore, there is a clear underinvestment in research and development in South Africa in general, not only in N&N research. The CSIR emerges as the most productive institute in N&N research, which can be attributed largely to the direct and continuous government investment in this organisation by, for example, establishing the National Centre for Nano-Structured Materials. However, innovative outputs, such as new nano-enabled products, as envisaged in the N&N strategy, will be difficult to achieve with a low level of research and collaboration in this field by the private sector. The introduction of R&D tax incentives in South Africa is a means to encourage an increased participation of the private sector in research and development. The lack of private sector involvement can in part be attributed to the early stage of nanotechnology development in South Africa with most research efforts geared to generating knowledge in the field. The more likely reason is the generally weak domestic corporate R&D expenditure, as currently the country's gross expenditure of R&D as a percentage of GDP of 0.73% is made up of only a small fraction from the corporate sector, as measured in 2014. In South Africa, the business expenditure of R&D as a percentage of GDP stands at 0.32% and this is against the OECD average of 1.58%, whereas that of China is 1.47%, showing that the country is lagging behind. Future work will need to look at a quantitative comparison of this output with other developing countries and possibly provide an analytical perspective of low patenting trends in this field.

Acknowledgement

We thank Sello Mutle for editing an earlier version of this manuscript.

Authors' contributions

X.M. undertook all the research and A.P. provided academic guidance.

References

1. South African Department of Science and Technology. South Africa's national research and development strategy. c2002 [cited 2016 Dec 09]. Available from: http://www.dst.gov.za/images/pdfs/National research development strategy 2002.pdf
2. South African Department of Science and Technology. Innovation towards a knowledge-based economy: Ten-year plan for South Africa 2008-2018 [document on the Internet]. c2007 [cited 2016 Dec 09]. Available from http://www.dst.gov.za/images/pdfs/The%20Ten-Year%20Plan%20for%20Science%20and%20Technology.pdf
3. Kaplan D. South Africa's national research and development strategy: A review. Sci Tech Soc. 2004;9(2):273–294. https://doi.org/10.1177/097172180400900204
4. Lux Research. Nanotechnology update: U.S. leads in government spending amidst increased spending across Asia. Boston, MA: Lux Research Inc.; 2015.
5. Reiss T, Thielmann A. Nanotechnology research in Russia: An analysis of scientific publications and patent applications. Nanotechnol Law Bus. 2010;7(4):387–404.
6. Organisation for Economic Co-operation and Development (OECD). OECD Nanotechnology in the context of technology convergence. c2014 [cited 2016 Dec 09]. Available from: http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=dsti/stp/nano(2013)10/final&doclanguage=en
7. South African Department of Science and Technology. National nanoscience and nanotechnology strategy. c2006 [cited 2016 Dec 09]. Available from: http://www.dst.gov.za/images/pdfs/Nanotech.pdf
8. Claassens CH, Motuku M. Nanoscience and nanotechnology research and development in South Africa. Nanotechnol Law Bus. 2006;3:217.
9. Pouris A. Nanoscale research in South Africa: A mapping exercise based on scientometrics. Scientometrics. 2007;70(3):541–553. https://doi.org/10.1007/s11192-007-0301-7
10. Ezema IC, Ogbobe PO, Omah AD. Initiatives and strategies for development of nanotechnology in nations: A lesson for Africa and other least developed countries. Nanoscale Res Lett. 2014;9(1), Art. #133, 8 pages. https://doi.org/10.1186/1556-276x-9-133
11. Maclurcan DC. Nanotechnology and developing countries: Part 2: What realities? Journal of Materials Online. 2005;1, 29 pages. http://dx.doi.org/10.2240/azojono0104
12. Shelton RD. Relations between national research investment and publication output: Application to an American paradox. Scientometrics. 2008;74(2):191–205. https://doi.org/10.1007/s11192-008-0212-2

13. Darvish H, Tonta Y. Diffusion of nanotechnology knowledge in Turkey and its network structure. Scientometrics. 2016;107(2):569–592. https://doi.org/10.1007/s11192-016-1854-0

14. Sooryamoorthy R. Science and scientific collaboration in South Africa: Apartheid and after. Scientometrics. 2010;84:373–390. https://doi.org/10.1007/s11192-009-0106-y

15. Kahn M. A bibliometric analysis of South Africa's scientific outputs – some trends and implications. S Afr J Sci. 2011;107(1/2), Art. #406, 6 pages. https://doi.org/10.4102/sajs.v107i1/2.406

16. Confraria H, Godinho MM. The impact of African science: A bibliometric analysis. Scientometrics. 2015;102(2):1241–1268. https://doi.org/10.1007/s11192-014-1463-8

17. Pouris A, Ho Y-S. Research emphasis and collaboration in Africa. Scientometrics. 2013;98(3):2169–2184. https://doi.org/10.1007/s11192-013-1156-8

18. Lavrik OL, Busygina TV, Shaburova NN, Zibareva IV. Nanoscience and nanotechnology in the Siberian Branch of the Russian Academy of Sciences: Bibliometric analysis and evaluation. J Nanopart Res. 2015;17(2), Art. #90, 11 pages. https://doi.org/10.1007/s11051-015-2900-1

19. Chiware E, Skelly L. Publishing patterns at the Cape Peninsula University of Technology. S Afr J Sci. 2016;112(1/2), Art. #2014-0220, 6 pages. https://doi.org/10.17159/sajs.2016/20140220

20. Leydesdorff L. The delineation of nanoscience and nanotechnology in terms of journals and patents: A most recent update. Scientometrics. 2008;76(1):159–167. https://doi.org/10.1007/s11192-007-1889-3

21. Garfield E. Science Citation Index: A new dimension in indexing. Science. 1964;144(3619):649–654. https://doi.org/10.1126/science.144.3619.649

22. Thomson Reuters. Web of Science™ report [created 2016 Feb 10].

23. Porter AL, Youtie J, Shapira P. Refining search terms for nanotechnology. J Nanopart Res. 2008;10(5):715–728. https://doi.org/10.1007/s11051-007-9266-y

24. Organisation for Economic Co-operation and Development (OECD). Science technology and patents [homepage on the Internet]. c2016 [cited 2016 Dec 09]. Available from: http://stats.oecd.org/

25. Maghreb M, Abbasi A, Amiri S, Monsefi R, Harati A. A collective and abridged lexical query for delineation of nanotechnology publications. Scientometrics. 2011;86:15–25. https://doi.org/10.1007/s11192-010-0304-7

26. Pouris A. A bibliometric assessment of energy research in South Africa. S Afr J Sci. 2016;112(11/12), Art. #2016-0054, 8 pages. http://dx.doi.org/10.17159/sajs.2016/20160054

27. Roco MC. The long view of nanotechnology development: The National Nanotechnology Initiative at 10 years. J Nanopart Res. 2011;13(2):427–445. https://doi.org/10.1007/s11051-010-0192-z

28. Chen H, Dang Y, Larson C, Roco MC, Wang X. Trends for nanotechnology development in China, Russia, and India. J Nanopart Res. 2009;11:1845–1866. https://doi.org/10.1007/s11051-009-9698-7

29. Appelbaum R, Parker R, Cao C. Developmental state and innovation: Nanotechnology in China. Glob Netw. 2011;11(3):298–314. https://doi.org/10.1111/j.1471-0374.2011.00327.x

30. Sargent JF. The National Nanotechnology Initiative: Overview, reauthorisation and appropriations issues [document on the Internet]. c2014 [cited 2016 Dec 09]. Available from: https://fas.org/sgp/crs/misc/RL34511.pdf

31. Frame DJ. Mainstream research in Latin America and the Caribbean. Interciencia. 1977;2:143–147.

32. Makhoba X, Pouris A. Scientometric assessment of selected R&D priority areas in South Africa: A comparison with other BRICS countries. Afr J Sci Technol Innov Develop. 2016;8(2):187–196. http://dx.doi.org/10.1080/20421338.2016.1147205

33. Finardi U. Scientific collaboration between BRICS countries. Scientometrics. 2015;102(2):1139–1166. https://doi.org/10.1007/s11192-014-1490-5

34. Boshoff N. South–South research collaboration of countries in the Southern African Development Community (SADC). Scientometrics. 2009;84(2):481–503. https://doi.org/10.1007/s11192-009-0120-0

Growth of soil algae and cyanobacteria on gold mine tailings material

AUTHORS:
Tanya Seiderer[1]
Arthurita Venter[1]
Fanus van Wyk[1]
Anatoliy Levanets[1]
Anine Jordaan[1]

AFFILIATION:
[1]Unit for Environmental Sciences and Management, North-West University, Potchefstroom, South Africa

CORRESPONDENCE TO:
Anatoliy Levanets

EMAIL:
20868421@nwu.ac.za

DATES:
Received: 09 Dec. 2016
Revised: 02 May 2017
Accepted: 20 June 2017

KEYWORDS:
revegetation; biological soil crusts; cyanobacteria; tailings storage facilities; scanning electron microscopy

HOW TO CITE:
Seiderer T, Venter A, Van Wyk F, Levanets A, Jordaan A. Growth of soil algae and cyanobacteria on gold mine tailings material. S Afr J Sci. 2017;113(11/12), Art. #2016-0384, 6 pages. http://dx.doi.org/10.17159/sajs.2017/20160384

ARTICLE INCLUDES:
× Supplementary material
× Data set

FUNDING:
None

The goal of revegetation of gold mine tailings storage facilities is to reduce aeolian pollution, nutrient leaching and erosion caused by exposure to wind and water. The establishment of biological soil crusts may prove to be a more cost-effective way to reach the same goal and the aim of this study was therefore to determine if it is possible to establish algae and cyanobacteria on gold mine tailings. Different treatments of *Chlamydomonas*, *Microcoleus* and *Nostoc* were inoculated on gold mine tailings in controlled conditions and algal growth was measured on all of the treatments after 6 weeks. *Nostoc* treatments had the highest chlorophyll-*a* concentrations and produced a surface crust, while *Chlamydomonas* treatments penetrated the tailings material and provided the strongest crust. The results were promising but more research is necessary to determine the best organism, or combination of organisms, to colonise mine tailings and to eventually produce biological crusts.

Significance:

- Determination of the best organisms to colonise mine tailings and to produce biological crusts for the revegetation of gold mine tailings storage facilities.

Introduction

Tailings storage facilities (TSFs) are hydraulically deposited mine residue storage facilities for potential harmful waste products such as waste rock, cyanided sand, slimes, surplus mine water and discarded solutions[1], also known as tailings material or mill tailings[2]. Uncovered slopes and sparse vegetation is a common sight on tailings material and these surfaces are exposed to the erosive forces of wind and water. Because of the often compacted nature of the material, decreased infiltration of water increases the possibility of run-off, erosion, sedimentation and air pollution through the generation of dust from the tailings material.[3] Dust emissions from the tailings may have health effects on communities in the vicinity of TSFs[1,4] and one of the goals of revegetation of these facilities is to reduce or eliminate human and wildlife exposure to contaminants that may be transported by airborne mine tailings[5]. The presence of acids and the high salt content of gold tailings material are likely synergistic in causing biotoxicity as low pH generally increases the bioavailability of metals.[6] As a result, most gold mine TSFs are devoid of vegetation and have a stressed heterotrophic microbial community.[5,7]

Rehabilitation methods implemented to reduce the environmental impact of erosion on TSFs include gravel mulching, rock cladding and vegetation establishment.[3,8] The plant canopy reduces aeolian dispersion while the plant roots prevent water erosion and nutrient leaching.[5] The establishment of vegetation for rehabilitation purposes is a very costly process because of the possible toxicity of heavy metals and immobilisation of some essential nutrients.[2] Therefore there is a great need for a cost-effective alternative to revegetate tailings material and provide reliable protection against erosion by wind and water.

Biological soil crusts (BSCs) consisting of bryophytes, lichens, algae and associated microorganisms are known to establish in extreme environments[9-11] such as mine tailings[12,13] and can contribute significantly to the stabilisation of soils through soil aggregate formation[14-17]. BSCs could therefore potentially provide protection against the erosive forces of wind and water on tailings material. Improved stability of soil inoculated with cyanobacteria has been identified[18] after only 6 weeks as a result of cyanobacteria binding and gluing aggregates and isolated mineral particles.

The aim of our study was to investigate establishment of algae and cyanobacteria on gold mine tailings in controlled conditions and to use scanning electron microscopy (SEM) to determine the success of colonisation. The first step was to establish a protocol for inoculation of algae and cyanobacteria on tailings material that can be used in the controlled study as well as in situ.

Materials and methods

Characterisation of the tailings material

Tailings material was collected from a gold mine TSF near Stilfontein (26.48° S, 26.47° E) in South Africa, and analysed in accordance with the standards set out by the Agricultural Laboratory Association of Southern Africa and the International Soil Analytical Exchange (Wageningen, the Netherlands).[19] Exchangeable Ca, Mg, K and Na were estimated by 1 M ammonium acetate (pH=7), P was estimated by P-Bray 1 extraction, pH was estimated via 1:2.5 extraction and conductivity was determined with a saturated extraction as described in NSSSA.[19]

Algal and cyanobacterial species present in the tailings material before inoculation and for all treatments following the 6-week trial period were determined by the methods described by Orlekowsky et al.[12]

Test organisms

The chlorophycean *Chlamydomonas* sp. and the cyanobacteria *Microcoleus vaginatus* and *Nostoc* sp. were isolated from the collected tailings material. The isolations were made by incubating tailings material on 1.5%

agarised Bold's basal growth medium (BBM)[20,21] in Petri plates at 20 °C in continuous light supplied from Philips Lifemax (TLD 58W/840) at 35 μmol photons/m^2/s (standard conditions). Following 14 days incubation, discrete colonies were transferred to fresh agarised medium in Petri plates in order to obtain unialgal cultures. Each species was then transferred to 100 mL BBM[21] and grown under standard conditions.

Cultures of *Chlamydomonas* sp., *M. vaginatus* and *Nostoc* sp. used for inoculation were grown in 200 mL BBM[21] under standard conditions until they reached the exponential growth phase.

Inoculation of tailings

The effectiveness of three inoculation methods was compared using *Chlamydomonas* sp. The three methods were:

1. A volume of 200 mL of *Chlamydomonas* sp. suspended in BBM was evenly poured onto the surface of tailings.
2. A volume of 200 mL *Chlamydomonas* sp. suspended in 1% agarised BBM was evenly poured on the surface of tailings.
3. A volume of 200 mL of *Chlamydomonas* sp. suspended in BBM was sprayed over the surface of tailings using a 2-L adjustable Thema spray bottle (Thema Home Kitchen Essentials).

Glasshouse trials

Rectangular seed trays (300 x 275 x 100 mm) were filled to a depth of 80 mm with gold mine tailings material and placed in a glasshouse in which the air temperature varied between 26 °C during the day and 20 °C at night. The light intensity ranged from 50 to 300 μmol photons/m^2/s depending on the time of day. Unsterilised tailings material was used because sterilisation by autoclaving resulted in excessive growth of fungi during the experiments.

Cultures grown in 200 mL BBM-agar suspension[21] were carefully poured over the tailings material. Considering high cost as a factor in tailings revegetation projects, cultures of algae and cyanobacteria were grown in both full strength and half strength BBM. The different treatments are shown in Table 1. There were three replicates of each.

Table 1: Treatments tested in the glasshouse

Treatment†	Inoculum	Water	Agarised Bold's basal medium		
			Full strength	Half strength PO_4^{3-}	Half strength NO_3^-
1	–	+			
2	–		+		
3	–			+	
4	–				+
5	*Chlamydomonas*		+		
6	*Microcoleus*		+		
7	*Nostoc*		+		
6	*Chlamydomonas*			+	
9	*Microcoleus*			+	
10	*Nostoc*			+	
11	*Chlamydomonas*				+
12	*Microcoleus*				+
13	*Nostoc*				+

†*Soils in all treatments contained the natural soil microflora.*

The trays were placed at random positions and watered for short intervals 5 times a day. After 3 weeks, 200 mL BBM growth medium (without any organism) was added to Treatments 2–13. Treatments 3 and 4 as well as Treatments 8–13 received 200 mL BBM with the same phosphate and nitrate concentrations that were initially added (see Table 1).

*Chlorophyll-*a *determination*

After 6 weeks, 10 g soil was collected to determine the chlorophyll-*a* concentration using the method described by Castle[22]. The soil sample in 9 mL methanol solution (2% ammonium acetate in 100% methanol) was placed on a shaker for 4 h, after which the sample was filtered through a Whatman GF/C filter. The absorbance of the supernatant was determined at 652 nm, 665 nm and 750 nm. To calculate the chlorophyll-*a* concentrations the following equation from Porra et al.[23] was used:

μg Chla/ g soil = [16.29*(A665 – A750) - 8.54*(A652 – A750)]*V)/g soil;

where V is the volume of solvent used (mL).

Penetration tests

Before inoculation, penetration tests were done on the dry tailings material with a handheld penetrometer that measures the force (in kg/cm^2) needed to break a soil crust. The penetration tests were repeated 3 weeks after the trial on dry soil. Three penetration tests were performed on each replicate of each treatment.

Scanning electron microscopy

After the 6-week period, three trays of each treatment were left to dry out for SEM analyses. An area of 1 cm^2 of the soil surface was sampled using a scalpel blade. Following the method of Tiedt et al.[24], the samples were put in 4% osmium vapour; dried and mounted on aluminium stubs using carbon tape and Leit C plus (Analytical Laboratory Solution). The mounted samples were sputter coated with gold-palladium and observed in a FEI Quanta 250 FEG scanning electron microscope at 8 kV.

Data analysis

The differences between the data sets were determined with Statistica version 13 software. The Kruskal–Wallis ANOVA was used (non-parametric data) for comparing multiple independent samples to determine differences between the different treatments. Standard error bars were also added to graphs (Figures 1 and 2) to indicate the variability of the data.

Results

The chemistry of the tailings material

The chemical analysis of the tailings material used is given in Table 2.

*Chlorophyll-*a *and penetration measurements*

There was a statistically significant difference between the chlorophyll-*a* of the control treatment that received only water (Treatment 1) and treatments *M. vaginatus* + agar (6), *Nostoc* sp. + agar (7), *M. vaginatus* + PO_4 (9), *Nostoc* sp. + PO_4 (10), *M. vaginatus* + NO_3 (12) and *Nostoc* sp. + NO_3 (13) (Table 3). The chlorophyll-*a* of Treatment 2 with BBM and agar and no organism also differed significantly from treatments *M. vaginatus* + agar (6), *Nostoc* sp. + agar (7), *M. vaginatus* + PO_4 (9), *Nostoc* sp. + PO_4 (10), *M. vaginatus* + NO_3 (12) and *Nostoc* sp. + NO_3 (13). Treatments 3 and 4 (both with no organisms) as well as Treatment 11 (*Chlamydomonas* sp. + NO_3) differed significantly from treatments *Nostoc* sp. + agar (7), *M. vaginatus* + PO_4 (9) and *M. vaginatus* + NO_3 (12) while Treatment 8 (*Chlamydomonas* sp. + PO_4) differed significantly from treatments *Nostoc* sp. + agar (7) and *M. vaginatus* + NO_3 (12) (Table 3). There was no statistically significant difference between the controls and *Chlamydomonas* sp. treatments (5, 8 and 11).

Table 2: Physical and chemical analysis of the gold mine tailings soil

Soil analyses	Gold mine tailings (<5% pyrite)
pH (H_2O)	5.4
EC (mS/m)	193
SO_4 (mg/kg)	1674
P (mg/kg)	1
K (mg/kg)	30
Ca (mg/kg)	1793
Mg (mg/kg)	94
Na (mg/kg)	9
CEC (cmol/kg)	1.3
Al (cmol/kg)	0.04
As (mg/kg)	0.05
Fe (mg/kg)	10.53
Mn (mg/kg)	14.90
Pb (mg/kg)	Not detected
Zn (mg/kg)	1.58

EC, electrical conductivity

All the *Nostoc* treatments grew well. Treatment 7 – the *Nostoc*–BBM–agar suspension (34.44 μg/g) – and Treatment 10 – the *Nostoc*–BBM–agar suspension with 8.1 g/L PO_4 (32.52 μg/g) – had the highest chlorophyll-*a* measurements (Figure 1). There was no statistical difference between these treatments and Treatment 13 in which *Nostoc* was grown in 9.12 g/L NO_3. Treatments with no test organism as well as treatments with *Chlamydomonas* sp. as test organism did not grow well.

The chlorophycean *Chlorococcum* sp. and species of the cyanobacteria *Nostoc*, *Phormidium* and *Scytonema* were present in the tailings material before inoculation. Table 4 shows that some organisms present in the tailings material before inoculation did not survive after the 6-week trial period. On the other hand, opportunistic genera such as *Arthrospira*, *Calothrix*, *Chlorolobion, Chlorosarcinopsis, Klebsormidium*, *Lyngbya*, *Navicula*, *Stichococcus* and *Tetracystis* were found after the 6-week trial period in different treatments.

Penetration was measured after the treatments were allowed to dry out. Tailings material with *Chlamydomonas* sp. cultured in BBM–agar suspension (Treatment 5) needed the greatest force to break the soil crust (2.58 kg/cm^2; Figure 2) but the force required did not differ significantly from that of the other treatments. The force required to penetrate the crust of the control differed significantly ($p=0.02$) from the force required to penetrate the crust of *Chlamydomonas* cultured in BBM–agar suspension with 8.1 g/L (Treatment 8).

Spraying delivered a more even coverage of the suspension but this method had a negative effect on the organisms and caused erratic growth. Pouring without agar showed a high initial growth but the chlorophyll-*a* concentrations (data not shown) declined from Day 6 onwards. The BBM–agar suspension that was poured on the surface proved to be the best method to inoculate algae and cyanobacteria into the soil.

Table 3: The *p*-values of chlorophyll-*a* measurements in comparisons of treatments showing a significant difference

Treatment	6	7	8	9	10	11	12	13
1	0.002	0.000001		0.000006	0.0001		0.00	0.003
2	0.009	0.000004		0.000031	0.0004		0.0000021	0.01
3		0.002		0.01			0.001	
4		0.003		0.01			0.002	
7			0.03			0.003		
8							0.02	
9						0.013		
11							0.0016	

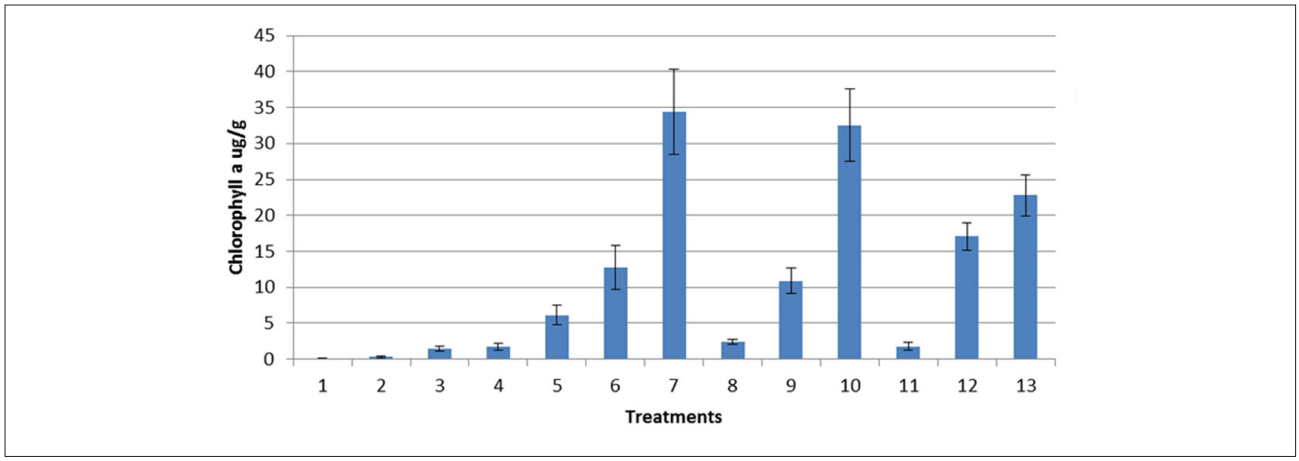


Figure 1: Mean chlorophyll-*a* concentrations in different treatments (1–13 listed in Table 1) following 6-week incubation.

Table 4: Genera present in the tailings material before inoculation and following 6-week treatment

Genus	Treatment													
	Before	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
Cyanobacteria														
Arthrospira							✓							
Calothrix		✓												
Lyngbya												✓		
Microcoleus							✓			✓			✓	
Nostoc	✓		✓	✓	✓	✓	✓	✓	✓		✓	✓		✓
Phormidium	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Scytonema	✓	✓												
Chlorophyta														
Chlamydomonas		✓		✓	✓	✓		✓	✓			✓		✓
Chlorella														✓
Chlorococcum	✓	✓		✓			✓				✓			
Chlorolobion											✓			
Chlorosarcinopsis					✓									
Klebsormidium			✓											
Stichococcus													✓	
Tetracystis				✓				✓	✓	✓				
Bacillariophyta														
Navicula												✓		
Total	**4**	**5**	**3**	**5**	**4**	**3**	**5**	**4**	**4**	**3**	**3**	**5**	**3**	**4**

See Table 1 for treatments.

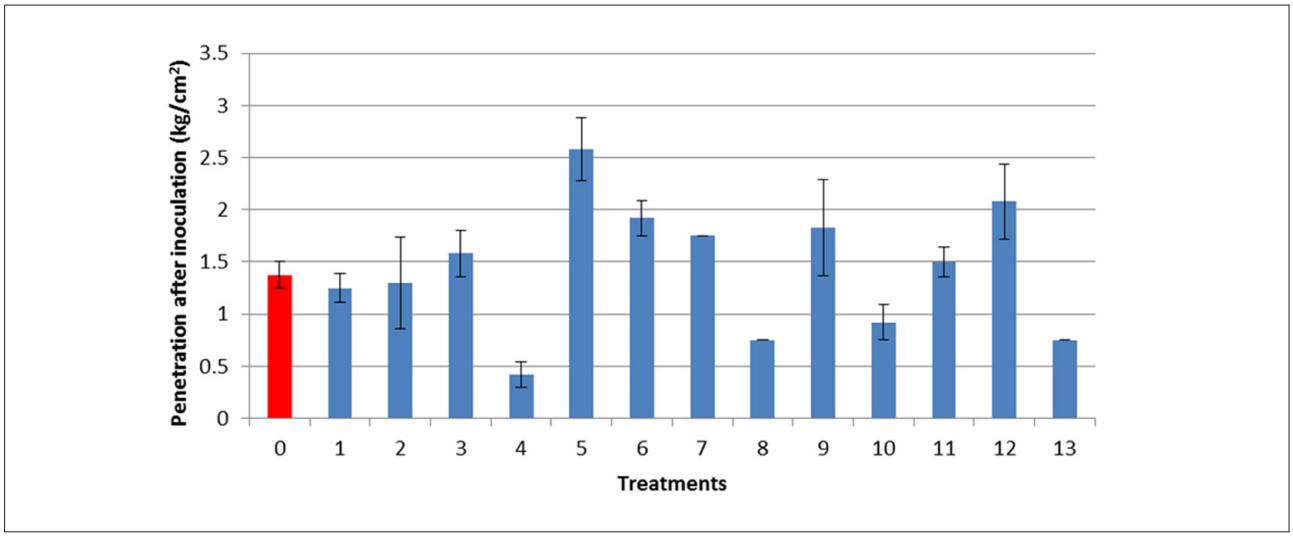


Figure 2: Results of soil penetration tests following treatment (see Table 1 for descriptions of treatments). The bar at 0 indicates penetration prior to inoculation.

This method showed a continuous increase in biomass, probably because the agar buffers the organism against the sudden change from a liquid medium to the harsher conditions of the soil. This method was therefore used in further trials.

Scanning electron microscopy

Figure 3 shows SEM images of the different treatments with the soil algae and cyanobacteria. *Phormidium* sp. grew on the surface of the tailings material – even in the case of the control treatment that received only water (Figure 3a). Crust development can be initiated and helped with input of water, nutrients and inoculum (Figure 3b).

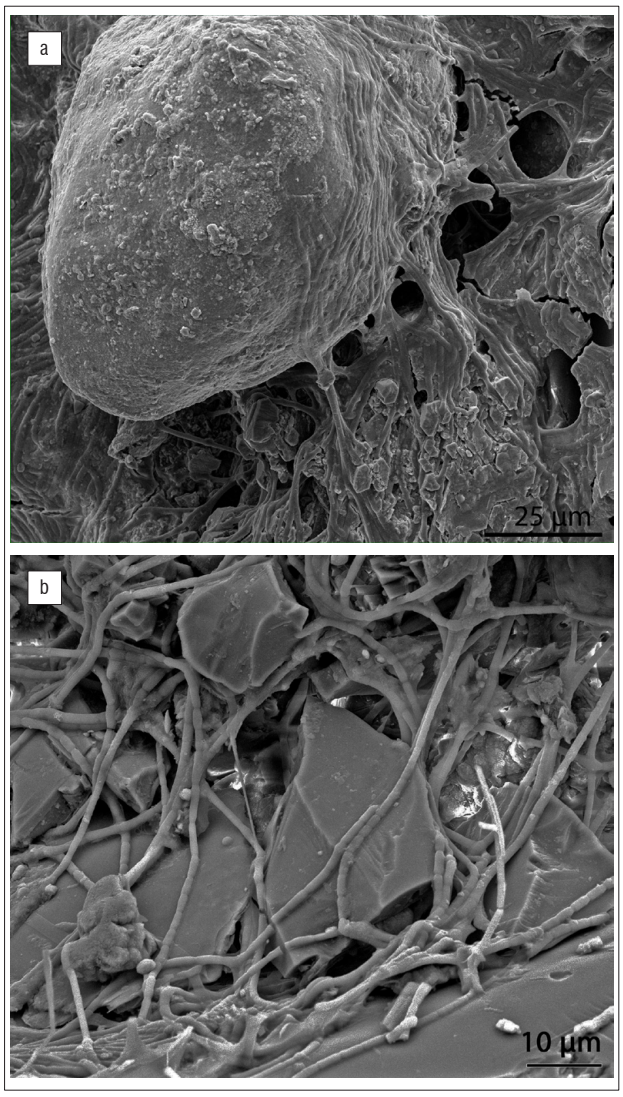


Figure 3: Scanning electron microscopy images showing the intimate relationship between algae, cyanobacteria and soil particles. (a) Control and (b) Treatment 9 after 6 weeks.

Discussion

The gold mine tailings material is acidic, has low cation exchange capabilities (CEC) and high electrical conductivity (EC; Table 2). According to Van Wyk[25], the ideal CEC for South African soil is about 5–20 cmol/kg and ideal EC between 60 and 100 mS/m. The CEC is the relative measure of the soil's ability to retain nutrients[26] and nutrients are less likely to leach out if CEC is high[27]. Phosphate levels are very low and although calcium levels are high, calcium may not be available for utilisation by the organisms because of the low pH.[28] The availability of nutrients to plants is determined by the oxidative state and chemical properties of the element while the soil pH, interactions with soil colloids, microbial activity and soil physical conditions such as aeration, compaction, temperature and moisture also play key roles.[29]

In spite of the harsh substrate conditions, this study shows that algae and cyanobacteria are able to colonise gold mine tailings and produce a viable crust within 6 weeks. There were significant differences between the treatments with *Nostoc* and *Microcoleus* and the treatments in which no organism was added (Table 3). The *Nostoc* treatments also had the highest chlorophyll-*a* concentrations and the lower phosphate and nitrate concentrations did not inhibit the growth of this organism (Figure 1; Treatments 10 and 13). *Nostoc* is a small and mostly immobile cyanobacterium that resides mostly on the soil surface and usually has a mucopolysaccharide sheath and UV-protective pigmentation to protect it from radiation exposure,[30] making it a good candidate to use as an inoculum on tailings material.

Better performance was expected from *Microcoleus* as BSC formation is often initiated by filamentous cyanobacteria such as *Microcoleus*[31] followed by smaller cyanobacteria and chlorophytes[32]. *Microcoleus* is also a common soil cyanobacterium[33] that occurred in most of the revegetated gold mine TSFs investigated by Orlekowsky et al.[12] This organism has the ability to move around within the soil. When the soil is wet they can move towards the surface into the photosynthetic zone, while they retract deeper into the soil when the soil dries out.[32] The continuous movement of the filaments in the soil ensures that sheath material is left behind in the soil layers thereby joining loose sand particles while forming a smooth biological soil crust on the soil surface. This type of crust may reduce water infiltration and can prevent the emergence of vascular plant seedlings[34] but may be a good cover to prevent dust emissions from mine tailings.

Of the three organisms chosen for this investigation, *Microcoleus* presented the most difficulties to work with as it produces films on the bottom of the culture flasks, making it difficult to spread the inoculum evenly on the tailings material, causing patchy growth patterns. However, treatments with *Microcoleus* (Treatments 6, 9 and 12) differed significantly from the control treatments (1, 2 in Table 3) and it is still a good candidate for inoculation on tailings material.

Chlorophytes such as *Chlamydomonas* are not pioneers and usually arrive later in[32] the development of BSCs. This late arrival may explain the low chlorophyll-*a* concentrations measured for the *Chlamydomonas* treatments. However, treatments with *Chlamydomonas* penetrated the tailings material up to 641.5 μm and it may help to stabilise the soil. Results indicated that *Chlamydomonas* Treatment 5 needed a force of 2.58 kg/cm^2 to break the crust. There was, however, no significant difference between the control and any of the treatments, emphasising that the crusts in all treatments were in a state of development.

The glasshouse trials were duplicated in field trials on TSFs with *Nostoc* and *Microcoleus* species (data not shown) and in spite of an untimely thunderstorm, preliminary results show a higher diversity in all the treatment plots after 6 weeks compared with those of the glasshouse trials, presumably as a result of the high diaspora concentration in the atmosphere surrounding the storage facilities.[12] However, the chlorophyll-*a* concentrations measured were significantly lower (<2 μg/g) than those measured in the glasshouse trials.

Conclusion

This study was a preliminary one to determine if algae and cyanobacteria can grow on mine tailings; we have shown that with the inoculation of algae and cyanobacteria, as well as the input of moisture and nutrients, biological crusts can be established on tailings material within 6 weeks. *Nostoc* treatments had the highest chlorophyll-*a* concentrations and produced a surface crust, while *Chlamydomonas* treatments penetrated the tailings material and provided a stronger crust. *Microcoleus* species are more difficult to work with but this species also produced a viable crust, making it a good candidate to use as inoculum. By improving growth conditions, diaspora from algae and cyanobacteria in the environment also germinated on the tailings material, increasing the probability of forming viable biological crusts. One such organism is the filamentous cyanobacterium *Phormidium* that occurred on all the treatments after 6 weeks and may also be a good candidate to use as an inoculant.

The results were promising but more research is necessary to determine the best organism, or combination of organisms, to colonise mine tailings and to eventually produce biological crusts. Moisture is very important and irrigation must be available when attempting field trials. The application of inoculum on a TSF must also be investigated as the conditions within a glasshouse are different from those of a TSF exposed to the outside environment.

Acknowledgements

We thank Eco-Analytica of the North-West University for performing the soil analyses.

Authors' contributions

F.v.W. conceptualised the project, acquired the funding and collected the data; T.S. undertook the data analysis and wrote the manuscript; T.S., A.V. and A.L. collected the data and analysed the samples; A.L. was responsible for student supervision; A.V. was the project leader, and was responsible for data curation and revisions; A.J. did the scanning electron microscopy.

References

1. Reichart M. A history of mine wastes rehabilitation techniques in South Africa: A multi-disciplinary overview of mine waste rehabilitation and the non-scientific drivers for its implementation 1950s – 1980s [PhD thesis]. Johannesburg: University of the Witwatersrand; 2012.
2. Haagner ASH. The role of vegetation in characterizing landscape function on rehabilitating gold tailings [MSc thesis]. Potchefstroom: North-West University; 2008.
3. Hattingh JM, Van Deventer PW. The effect of the chemical properties of tailings and water application on the establishment of a vegetative cover on gold tailings dams. Water Research Commission report no. 899/1/04. Pretoria: Water Research Commission; 2004.
4. Smallhorne M. Joburg's iconic mine dumps are a health risk, say activists. Mail and Guardian. 2012 December 14; Arts and culture.
5. Mendez MO, Maier RM. Phytoremediation of mine tailings in temperate and arid environments. Rev Environ Sci Biotechnol. 2008;7(1):47–59. http://dx.doi.org/10.1007/s11157-007-9125-4
6. Stevenson FJ, Cole MA. Cycles of soil: Carbon, nitrogen, phosphorus, sulphur, micronutrients. New York: John Wiley and Sons Inc.; 1999.
7. Moynahan OS, Zabinski CA, Gannon JE. Microbial community structure and carbon-utilization diversity in mine tailings revegetation study. Restor Ecol. 2002;10:77–87. http://dx.doi.org/10.1046/j.1526-100X.2002.10108.x
8. Blight G. Mine waste: A brief overview of origins, quantities, and methods of storage. In: Letcher T, Vallero D, editors. Waste: A handbook for management. Cambridge: Academic Press; 2011. p. 77–88. https://doi.org/10.1016/B978-0-12-381475-3.10005-1
9. Shields LM, Durell LW. Algae in relation to soil fertility. Bot Rev. 1964;30(1):92–128. http://dx.doi.org/10.1007/BF02858614
10. Eldridge DJ, Zaady E, Shachak M. Infiltration through three contrasting biological soil crusts in patterned landscapes in the Negev, Israel. Catena. 2000;40:322–336. http://dx.doi.org/10.1016/S0341-8162(00)00082-5
11. Venter A, Levanets A, Siebert S, Rajakaruna N. A preliminary survey of the diversity of soil algae and cyanoprokaryotes on mafic and ultramafic substrates in South Africa. Austral J Bot. 2015;63:341–352. http://dx.doi.org/10.1071/BT14207
12. Orlekowsky T, Venter A, Van Wyk F, Levanets A. Cyanobacteria and algae of gold mine tailings in the Northwest Province of South Africa. Nov Hedw. 2013;97(3–4):281–294. http://dx.doi.org/10.1127/0029-5035/2013/0117
13. Frouz J, Keplin B, Pizl V, Tajovsky K, Stary J, Lukesova A, et al. Soil biota and upper soil layer development in two contrasting post-mining chronosequences. Ecol Eng. 2001;17:275–284. http://dx.doi.org/10.1016/S0925-8574(00)00144-0
14. Lange OL. Photosynthesis of soil-crust biota as dependent on environmental factor. In: Belnap J, Lange OL, editors. Biological soil crusts: Structure, function, and management. Berlin: Springer; 2001. p. 217–240. https://doi.org/10.1007/978-3-642-56475-8_18
15. Hu C, Zhang D, Huang Z, Liu Y. The vertical microdistribution of cyanobacterial and green algae within desert crusts and the development of the algal crusts. Plant Soil. 2003;257:97–111. https://doi.org/10.1023/A:1026253307432
16. Flechtner V. North American desert microbiotic soil crust communities: Diversity despite challenge. In: Seckbach J, editor. Algae and cyanobacteria in extreme environments. Dordrecht: Springer; 2007. p. 539–551. https://doi.org/10.1007/978-1-4020-6112-7_29
17. Bowker MA. Biological soil crust rehabilitation in theory and practice: An underexploited opportunity. Restor Ecol. 2007;15(1):13–23. http://dx.doi.org/10.1111/j.1526-100X.2006.00185.x
18. Issa OM, Défarge C, Bissonnais YL, Marin B, Duval O, Bruand A, et al. Effects of the inoculation of cyanobacteria on the microstructure and the structural stability of a tropical soil. Plant Soil. 2007;290(1–2):209–219. http://dx.doi.org/10.1007/s11104-006-9153-9
19. Non-Affiliated Soil Analysis Work Committee, Soil Science Society of South Africa. Handbook of standard soil testing methods for advisory purposes. Pretoria: Soil Science Society of South Africa; 1990.
20. Brown RM, Larson DA, Bold HC. Airborne algae: Their abundance and heterogeneity. Science. 1964;143(3606):583–585. http://dx.doi.org/10.1126/science.143.3606.583
21. Stein JR. Handbook of phycological methods and culture methods and growth measurements. Cambridge: Cambridge University Press; 1973.
22. Castle SC. Chlorophyll-*a* double extraction with methanol. Boulder, CO: Aridlands Ecology Laboratory, University of Colorado; 2010.
23. Porra RJ, Thompson WA, Kriedeman PE. Determination of accurate extinction coefficients and simultaneous equations for assaying chlorophylls *a* and *b* extracted with four different solvents: Verification of the concentration of chlorophyll standards by atomic spectroscopy. Biochim Biophys Acta. 1989;975:384–394. https://doi.org/10.1016/S0005-2728(89)80347-0
24. Tiedt LR, Jooste WJ, Hamilton-Attwell VL. Technique for preserving aerial fungus structures for scanning electron microscopy. Trans Br Mycol Soc. 1987;88(3):420–422. https://doi.org/10.1016/S0007-1536(87)80021-9
25. Van Wyk SJ. An analytical investigation of the biophysical factors that inhibit successful ecological restoration of gold tailings dams [MEnvSci thesis]. Potchefstroom: North-West University; 2002.
26. Winegardner DL. An introduction to soils for environmental professionals. Boca Raton, FL: CRC/Lewis Publishers; 1995.
27. Hardy DH, Tucker MR, Messick JK, Stokes C. Understanding the soil test report. Raleigh, NC: Agronomic Division, North Carolina Department of Agriculture & Consumer Services; 2012.
28. Miles N, Farina M. Soil acidity and its management in crop production [homepage on the Internet]. c2013 [cited 2016 Nov 27]. Available from: http://www.grainsa.co.za/soil-acidity-and-its-management-in-crop-production
29. Hodges SC. Soil fertility basics: NC Certified Crop Advisor Training. Raleigh, NC: Soil Science Extension, North Carolina State University; 2010.
30. Belnap J, Phillips SL, Miller ME. Response of desert biological soil crusts to alterations in precipitation frequency. Oecologia. 2004;141:306–316. https://doi.org/10.1007/s00442-003-1438-6
31. Mager DM, Hui CA. A first record of biological crusts in the Cape Floristic Region. S Afr J Sci. 2012;108(7/8), Art. #1013, 4 pages. http://dx.doi.org/10.4102/sajs.v108i7/8.1013
32. Belnap J. Factors influencing nitrogen fixation and nitrogen release in biological soil crusts. In: Belnap J, Lange OL, editors. Biological soil crusts: Structure, function, and management. Berlin: Springer; 2003. p. 241–261. https://doi.org/10.1007/978-3-642-56475-8
33. Metting B. The systematics and ecology of soil algae. Bot Rev. 1981;47(2):195–312. http://dx.doi.org/10.1007/BF02868854
34. Belnap J, Eldridge D, Hilty-Kaltenecker J, Leonard S, Rosentreter R. Biological soil crusts: Ecology and management. Technical reference 1730-2. Denver, CO: United States Department of the Interior Bureau of Land Management; 2001.

Forecasting winter wheat yields using MODIS NDVI data for the Central Free State region

AUTHORS:
Zinhle Mashaba[1,2]
George Chirima[1,3]
Joel O. Botai[2,4]
Ludwig Combrinck[2]
Cilence Munghemezulu[2]
Ernest Dube[5]

AFFILIATIONS:
[1]Agricultural Research Council – Institute for Soil, Climate and Water, Pretoria, South Africa
[2]Department of Geography, Geoinformatics and Meteorology, University of Pretoria, Pretoria, South Africa
[3]School of Animal, Plant and Environmental Sciences, University of the Witwatersrand, Johannesburg, South Africa
[4]South African Weather Service, Pretoria, South Africa
[5]Agricultural Research Council – Small Grain Institute Production Systems, Bethlehem, South Africa

CORRESPONDENCE TO:
Zinhle Mashaba

EMAIL:
MashabaZ@arc.agric.za

DATES:
Received: 05 July 2016
Revised: 06 Feb. 2017
Accepted: 20 June 2017

KEYWORDS:
dryland wheat; wheat yield; NDVI; MODIS; food security

HOW TO CITE:
Mashaba Z, Chirima G, Botai JO, Combrinck L, Munghemezulu C, Dube E. Forecasting winter wheat yields using MODIS NDVI data for the Central Free State region. S Afr J Sci. 2017;113(11/12), Art. #2016-0201, 6 pages. http://dx.doi.org/10.17159/sajs.2017/20160201

ARTICLE INCLUDES:
× Supplementary material
× Data set

FUNDING:
Agricultural Research Council; National Research Foundation (South Africa); University of Pretoria

Consumption of wheat is widespread and increasing in South Africa. However, global wheat production is projected to decline. Wheat yield forecasting is therefore crucial for ensuring food security for the country. The objective of this study was to investigate whether the anthesis wheat growth stage is suitable for forecasting dryland wheat yields in the Central Free State region using satellite imagery and linear predictive modelling. A period of 10 years of Normalized Difference Vegetation Index data smoothed with a Savitzky–Golay filter and 10 years of wheat yield data were used for model calibration. Diagnostic plots and statistical procedures were used for model validation and assessment of model adequacy. The period 30 days before harvest during the anthesis stage was established to be the best period during which to use the linear regression model. The calibrated model had a coefficient of determination of 0.73, a *p*-value of 0.00161 and a root mean squared error of 0.41 tons/ha. Residual plots confirmed that a linear model had a good fit for the data. The quantile-quantile plot provided evidence that the residuals were normally distributed, which means that assumptions of linear regression were fulfilled and the model can be used as a forecasting tool. Model validation showed high levels of accuracy. The evidence indicates that use of Moderate Resolution Imaging Spectroradiometer data during the anthesis growth stage is a reliable, cost-effective and potentially time-saving alternative to ground-based surveys when forecasting dryland wheat yields in the Central Free State.

Significance:

- Developing a cost-effective technique based on satellite imagery for wheat yield forecasting is vital for food security planning in South Africa.

Introduction

Wheat (*Triticum aestivum* L.) is an important crop in many parts of the world including South Africa, where it is the second largest component of the staple diet after maize.[1,2] Consequently, it is crucial to predict wheat yields as global wheat production is expected to decrease under conventional management as a result of climate variability.[3,4] Additionally, a challenge exists to feed a growing human population while avoiding environmental problems such as deforestation and land degradation.[5] The central Free State Province of South Africa is a land-locked, dryland wheat production region, which exhibits variable agricultural production as a consequence of droughts and a reduced capacity to operate in world markets owing to high transport costs and foreign exchange constraints.[6,7] In order to ensure food security, there is a need for generating timely and accurate information on crop yields.[8] We report here on the development of a reliable estimate of wheat yields using the Moderate Resolution Imaging Spectroradiometer – Normalized Difference Vegetation Index (MODIS-NDVI). Accurate forecasting of the yield potential of dryland wheat in the Central Free State region will aid agricultural decision-makers in balancing the trade of agricultural commodities and reducing short-term price instabilities.[9]

Commonly, yield forecasting models are more reliable if applied during specific critical plant growth stages. For wheat, the anthesis (or flowering) stage appears to be an important stage.[10-12] During this time, water deficiencies lead to yield losses by reducing the spike and spikelet numbers, as well as the fertility of the remaining spikelets.[13] Water shortages during this stage also accelerate leaf senescence and reduce the rate of grain filling, thereby reducing the mean kernel weight.[14] High temperatures reduce the number of grains per ear, kernel weight and harvest index, leading to reduced grain yields.[15]

Vegetation indices derived from remote sensing technology are often used for crop monitoring and crop yield estimates.[2] The technique is based on the assumption that spectral data are related to canopy reflectance parameters which, in turn, are related to the final yield.[16,17] The NDVI used in this study can be used as an indicator of the photosynthetic potential of a vegetation canopy.[18] The NDVI makes use of the near infrared band and visible red band in the electromagnetic spectrum.[19] The limitation of this index is that it gets saturated in areas with dense biomass, but hyperspectral data can overcome this limitation.[20,21] Hyperspectral imaging is, however, costly as it requires a dedicated campaign, has a limited extent and a complex data structure compared to MODIS data, which is freely available.

Numerous studies have used NDVI derived from various remotely sensed images for crop yield estimations.[22-24] For example, Yang et al.[22] evaluated the accuracy of QuickBird satellite imagery and airborne imagery for mapping grain sorghum yield patterns. The results illustrated that QuickBird and airborne imagery had similar correlations with grain yield at 2.8-m and 8.4-m resolutions. Nuarsa et al.[23] estimated rice yields using an exponential model derived from Landsat Enhanced Thematic Mapper plus (ETM+) NDVI and field-observed rice yield in the Tabanan Regency. The study observed a coefficient of determination (R^2) and standard error of 0.852 and 0.077 tons/ha, respectively. Mutanga et al.[24] determined the optimal time for predicting sugarcane yield to be 2 months before harvest using NDVI derived from SPOT images.

However, few studies have been done on using remote sensing data for yield estimation in South Africa.[25,26] For example, Unganai and Kogan[25] demonstrated that corn yields can be estimated using Advanced Very High Resolution Radiometer (AVHRR) data with a spectral resolution of 0.58–12.5 μm. In the study it was found that Vegetation Condition Index and Temperature Condition Index derived from AVHRR data are highly correlated with corn yields. Frost et al.[26] demonstrated that Terra MODIS (0.6–1.1 μm) satellite sensor data products can be applied for maize yield estimation in South Africa. In that study, the window method was used and the resulting window periods showed that average NDVI and average Enhanced Vegetation Index data can be used for maize yield estimations.

Further studies are needed in South Africa on using remote sensing for yield predictions. This topic is important because wheat production is decreasing as a result of weather variability within the summer rainfall region, the deregulation of the wheat industry, and farmers converting to sustainable crops (e.g. soybean and canola).[1] Furthermore, timeous generation of yield projections will support timeous decisions concerning either importation or exportation of wheat. Therefore, the overall objective of this study was to develop a yield model for dryland winter wheat for the Central Free State region using MODIS data. This objective was undertaken by investigating whether wheat yields in that region are correlated with the MODIS-NDVI during the anthesis stage[10-12], and validating the performance and adequacy of the calibrated model.

Data and methods

Study area

The Free State Province hosts four distinct dryland wheat production regions: Central Free State, North Western Free State, South Western Free State and Eastern Free State.[27,28] These regions receive summer rainfall and experience frequent droughts. Therefore, farmers adopt farm management practices which make efficient use of rain for crop production. The underlying geology of the Free State is rocks from the Beaufort and Ecca Groups of the Karoo Supergroup, which make up the parent material for the soils.[27] The study sites of Arlington, Tweespruit and Excelsior are in the Central Free State wheat production region (Figure 1). These sites are part of the National Wheat Cultivar Evaluation Programme conducted by the Small Grain Institute of the Agricultural Research Council (ARC-SGI). The programme delivers information about the performance of wheat varieties from the major breeding companies of South Africa. The sites were selected systematically in a manner such that they are representative of all the production conditions of this geographical region.[29] Dryland wheat planting normally takes place from the first week of July (South African winter).[28]

MODIS-NDVI

The 16-day composite MODIS-NDVI (MOD13Q1) product images with a 250-m resolution used in this study are freely available from the US National Aeronautical Space Agency (NASA) Earth Observing System (EOS) website. The data obtained were for a 10-year period from 2000 to 2013 (excluding years 2001, 2002, 2008 and 2011) based on the available wheat yield data collected at the study sites. The MOD13Q1 product is computed from the surface reflectance of each band (red and near-infrared) as it would have been measured at ground level if there was no atmospheric scattering or absorption.[30] During data processing, corrections are made for the effects of atmospheric gases, aerosols and thin cirrus clouds.[31]

The downloaded MODIS-NDVI images were reprojected from Sinusoidal Projection to Geographic Projection using the MODIS Reprojection Tool.[32] Additionally, rescaling of the raster images had to be done to correct the range of NDVI values to range from -1 to 1. The raster images were cropped with the wheat boundaries obtained from the ARC in collaboration with Geo Terra Image and Spatial Business Intelligence. The values for NDVI were extracted for each of the wheat boundary pixels at the three localities.

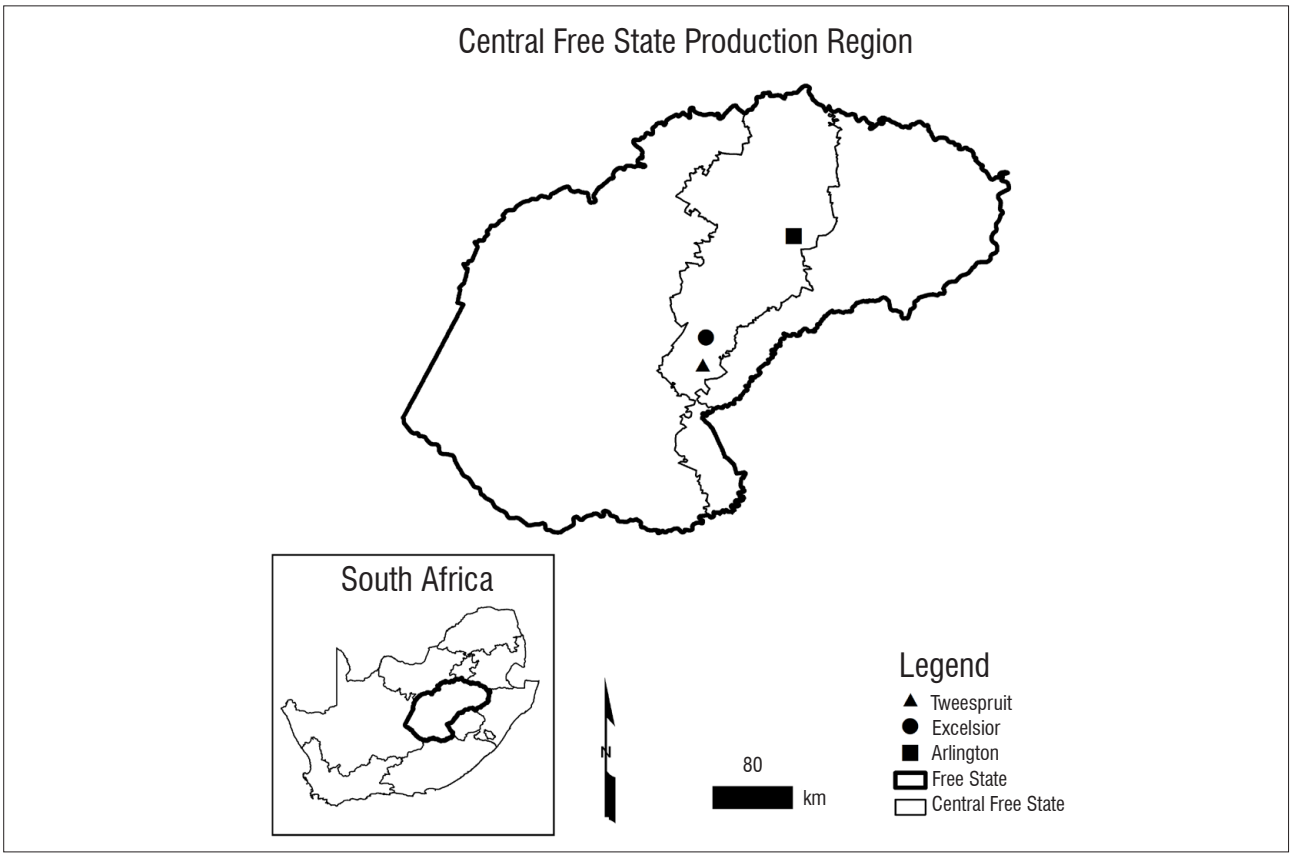


Figure 1: Map of the Central Free State wheat production region depicting the three study sites.

Wheat yield data

Annual wheat yield data for 10 seasons (2000–2013 excluding 2001, 2002, 2008 and 2011) were used for this study. These data were collected by the ARC-SGI under the National Wheat Cultivar Evaluation Programme. A randomised complete block design was used for the entire trials layout. The trials were planted inside non-irrigated wheat fields, in line with crop management practices with regard to tillage practices, seeding rates, weed control, fertiliser application, pest and disease control as well as planting date. The dryland wheat trials consist of five rows that are each 5 m long with an inter-row spacing of 0.45 m. The harvest area is 5 m × 1.35 m, represented by three central rows. In order to adequately test the cultivars on the planting date spectrum, which is as long as 3 months, two independent randomised trials (early and late planting) were planted at all sites within the geographical regions.

Savitzky–Golay filter applied to NDVI time series

The Savitzky–Golay[33] (S-G) filter algorithm was used to smooth out the MODIS-NDVI data. The S-G filter is based on local least-squares polynomial approximation. The advantages of the S-G filter are that it preserves features of the data such as relative maxima, minima and widths.[34] The S-G smoothing algorithm is given by Equation 1:

$$Y_j^* = \frac{\sum_{i=-m}^{i=m} C_i Y_{j+i}}{N}, \qquad \text{Equation 1}$$

where Y_j^* is the filtered value, C_i is the coefficient for the i-th NDVI value of the filter, Y_{j+i} represents the original NDVI value, and N is the number of convoluting integers equal to smoothing window size $(2m+1)$.[33] The larger the value of m, the smoother the results at the expense of flattening sharp peaks.[34]

Model development

A linear regression model was developed between the average yield of different late planted wheat cultivars and the average NDVI for the three study sites. The average yield was considered to be an independent variable and the average NDVI was considered to be a dependent variable according to:

$$P(Y|x) = \beta_0 x + \beta_1, \qquad \text{Equation 2}$$

where $P(Y|x)$ is the predicted yield as function of NDVI, x is the NDVI, β_0 is the coefficient and β_1 is the constant for winter wheat yield. Different models (logarithmic, exponential and power) were compared for the purpose of evaluating the models so that the best fitting model was selected.[35] Other studies[2,12,36] have found the linear model to be ideal for winter wheat yield and NDVI in various regions.

Model validation

Statistical tests were performed to validate the performance of the model. The goodness of fit of the model and the percentage of variance explained by the model were assessed using the coefficient of determination. The significance of the model was tested by means of p-values. The root mean squared error (RMSE) was also included in the analysis. Diagnostic plots were constructed to compare the observed yield and the predicted yield. The residuals were plotted against the wheat yield, which was necessary to check for linearity of the data and the presence or absence of inhomogeneity of variance.[37] Additionally, a quantile-quantile (Q-Q) residual plot was used to assess how close the theoretical distribution was to the model distribution. A normal distribution is indicated by a strong linear pattern for the sample points, and outliers can also be detected by visual inspection of the plot.[38] These validation methods differ from the widely used methods, which mostly focus on the RMSE and comparing the correlation between observed yield and predicted yield. The validation techniques also aided in understanding the underlying trends in the data.

Model testing

The NDVI data for the years 2001, 2002, 2008, 2011 and 2014 were used for model testing. These years were not used in model calibration and model validation. The NDVI of these years was used to predict the expected wheat yield. The predicted yield was then subtracted from the observed yield and the percentage error was calculated for the observed yield. The standard error was used as a measure of the accuracy of the predicted yield for which values close to zero indicate high accuracy. The year 2015 could not be used in the analysis because of severe drought in the non-irrigated wheat regions.

Results and discussion

Relationships between wheat yield and NDVI data

The best fitting model between wheat yield and NDVI was developed by using the Zadoks scale[39] to identify the critical growth stages of wheat. According to previous research carried out in other regions, the most critical growth stage for dryland wheat is anthesis.[10-12] This stage was also highly correlated with the final wheat yield in the Central Free State, as with other regions. The linear relationship between the average yield and average NDVI is represented by Equation 3:

$$P(Y|x) = 12.1136x - 2.7307 \ (R^2 = 0.73, p = 0.00161) \qquad \text{Equation 3}$$

In this study, the seasonal maximum NDVI was used to correlate with average wheat yields. This time occurs approximately 30 days prior to harvest during the first week of November (day of year 305 in regular years). The range of NDVI values for the model is 0.32–0.49 (Figure 2). These values fall within the threshold indicated by Ren et al.[2] for winter wheat of 0.2–0.8. The model was calibrated using NDVI but other parameters such as growing degree days or heat units, soil conditions and weather conditions can also be considered in order to improve the accuracy of the model.

Model validation

The regression models' predicted yield was compared with the observed yield from the 10-year winter wheat yield data (Figure 2). The p-value was 0.00161 ($p < 0.05$) and the R^2 value was 0.73 between wheat yield and NDVI, indicating a good relationship between the variables. These results are similar to those reported by Lopresti[12] of an R^2 of 0.75 for winter wheat yield in Northern Buenos Aires Province, Argentina. The similarity of these results could be because both Argentina and South Africa are in the southern hemisphere. Periods for winter wheat production are similar for both countries because the seasonal cycles (winter from June to September) coincide. However, Ren et al.[2] observed an R^2 of 0.88 for Shandong (China) when relating the production of winter wheat with the accumulated MODIS-derived NDVI.

The RMSE of the calibrated model was validated against the observed yield. An RMSE of 0.41 tons/ha results from using the least squares regression line to predict the wheat yield. Becker-Reshef et al.[8] reported a similar RMSE of 0.44 tons/ha for Kansas using MODIS-NDVI for wheat forecasting. Moriondo et al.[11] observed an RMSE of 0.44 tons/ha and 0.47 tons/ha, respectively, when using AVHRR-NDVI data to develop a wheat yield model for two Italian provinces.

Diagnostic plots were constructed to assess the fit of the model and whether the residuals are normally distributed. Residual plots (Figure 3a) were used to determine if the model was linear. Linear models have a random scatter of data points whereas non-linear models have a distinctive pattern.[37] A random dispersal of the residuals was observed, which means that the linear model is an ideal fit between the wheat yield and NDVI. The Q-Q plot depicted in Figure 3b indicated that the residuals are homogeneous, although small variations were present at the lower and upper tails of the plot. Periods of drought could induce such variations (outliers) and reduce the wheat yield as dryland wheat relies on residual soil moisture for growth. Overall, the residuals are normally distributed as they lie close to a straight line.

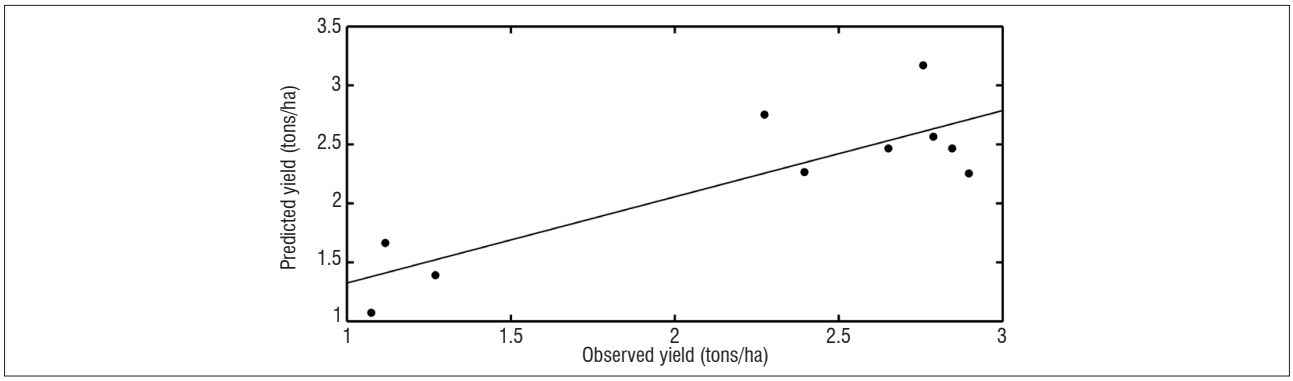


Figure 2: Central Free State predicted wheat yield as a function of observed wheat yield.

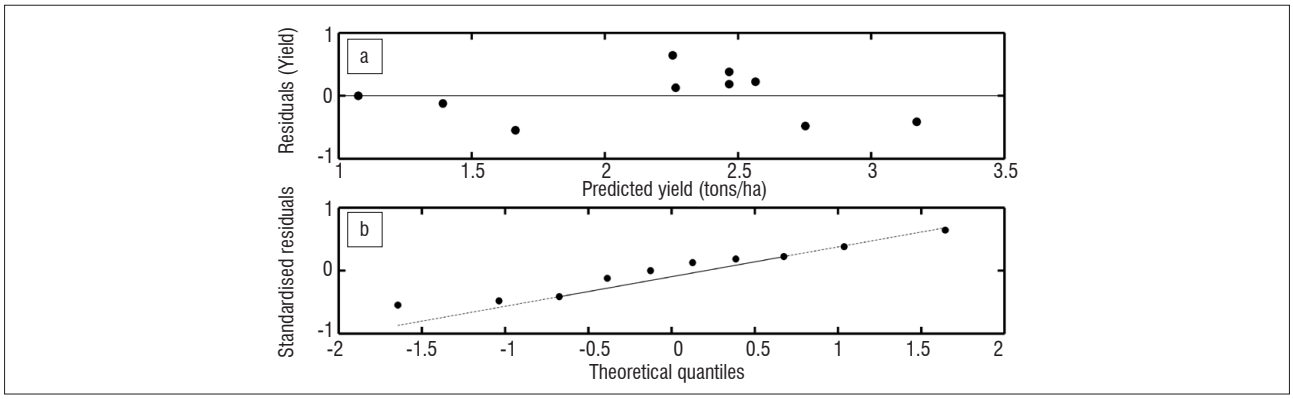


Figure 3: Model validation diagnostic plots. (a) Residual plot displaying randomly distributed residuals indicating that the linear model is the appropriate fit for the data. (b) Q-Q plot indicating that the residuals are normally distributed.

Table 1: Model testing results for the observed and predicted yield

Year	Observed yield (O) (tons/ha)	Predicted yield (P) (tons/ha)	Difference (O-P) (tons/ha)	Difference for observed yield (%)	Standard error (tons/ha)
2001	2.7700	2.9627	-0.1927	-6.9567	0.0862
2002	1.9900	1.2300	0.7600	38.1910	0.3399
2008	1.2400	1.0003	0.2397	19.3306	0.1072
2011	1.7100	1.1730	0.5370	31.4035	0.2402
2014	2.4000	2.1300	0.2700	11.2500	0.1207

These diagnostic plots were necessary as other studies have observed different relationships between vegetation indices and crop yields. Hayes and Decker[40] observed a quadratic relationship between a satellite data variable, Vegetation Condition Index, Crop Moisture Index and a climatological variable for predicting maize production. The R^2-value of the model was 0.73. Ma et al.[41] reported a power function to be representative of the relationship between soybean yield and canopy reflectance measurements at different soil types. The R^2-value of the model was 0.80. Benedetti and Rossini[42], Ren et al.[2] and Lopresti et al.[12] all observed a linear relationship between wheat yield and NDVI. However, Mkhabela et al.[43] observed that a power function was representative of the relationship between NDVI and spring wheat in different environments. The differences in models for spring wheat and winter wheat could be induced by different weather conditions and irrigation as winter wheat in South Africa is not irrigated.

Model testing

The percentage errors in Table 1 indicate that the model slightly overestimated yield in 2002, 2008 and 2011 because the values are above 10%. However, the model performed well in 2001 as the percentage errors are within the threshold of ±10%. The model predicted wheat yield well in 2014 as this year did not fall within the 2000–2013 period for which the model was calibrated. The standard errors were close to zero, which means that the predicted yield has a reasonable level of accuracy and, thus, the model was reliable.

Throughout this study MODIS data from an optical sensor was used but synthetic aperture radar (SAR) data can also be used. The advantages of SAR are that it can operate during the day or night, and in rainy or cloudy conditions. However, SAR data require complex processing, by often specialised software such as the INAHOR[44] and thus were not a focus of

this study. Data provided by satellites at a high resolution, such as SPOT or RapidEye, are better for small-scale estimates of crop yield because many tiles are needed to cover a large area, which also is costly.

Conclusion

The prospect of using MODIS-NDVI for winter wheat yield forecasts in the Central Free State production region was investigated using regression models. Findings suggest the best time to relate MODIS-NDVI to final wheat yields for this area is the period leading to 30 days before harvest (first week of November). This period coincides with the anthesis stage, and at this time, wheat yield is highly correlated with NDVI. The relationship between NDVI and wheat yield was significant with an R^2-value of 0.73, a p-value of 0.00161 and an RMSE of 0.41 tons/ha. Furthermore, diagnostic plots, model testing and validation provided evidence of the reasonable levels of model accuracy, model reliability and a good fit. These techniques complement the widely used techniques of comparing the correlation between the observed yields and predicted yields, and using the RMSE.

Acknowledgements

We thank the Agricultural Research Council, the National Research Foundation of South Africa and the University of Pretoria for funding. We thank the various collaborators (Geo Terra Image, Spatial Business Intelligence and the National Wheat Cultivar Evaluation Programme) for help in collecting the data on wheat farm boundaries and wheat yields. We also thank the Centre for Geoinformation Science of the University of Pretoria for their continued support.

Authors' contributions

Z.M. conceptualised and designed the study and wrote the first draft of the manuscript; G.C. was the project leader and budget owner; G.C., L.C. and J.O.B. modified, edited and approved the manuscript; C.M. was involved in data analysis; and E.D. edited the manuscript and was involved in field data collection.

References

1. Breitenbach MC, Fenyes TI. Maize and wheat production trends in South Africa in a deregulated environment. Agrekon. 2000;39(3):292–312. https://doi.org/10.1080/03031853.2000.9523578
2. Ren J, Chen Z, Zhou Q, Tang H. Regional yield estimation for winter wheat with MODIS-NDVI data in Shandong, China. Int J Appl Earth Obs Geoinf. 2008;10(4):403–413. https://doi.org/10.1016/j.jag.2007.11.003
3. Parry ML, Rosenzweig C, Iglesias A, Livermore M, Fischer G. Effects of climate change on global food production under SRES emissions and socio-economic scenarios. Glob Environ Change. 2004;14(1):53–67. https://doi.org/10.1016/j.gloenvcha.2003.10.008
4. Ortiz R, Sayre KD, Govaerts B, Gupta R, Subbarao GV, Ban T, et al. Climate change: Can wheat beat the heat?. Agric Ecosyst Environ. 2008;126(1):46–58. https://doi.org/10.1016/j.agee.2008.01.019
5. Balkovič J, Van der Velde M, Skalský R, Xiong W, Folberth C, Khabarov N, et al. Global wheat production potentials and management flexibility under the representative concentration pathways. Glob Planet Change. 2014;122:107–121. https://doi.org/10.1016/j.gloplacha.2014.08.010
6. Hammer GL, Hansen JW, Phillips JG, Mjelde JW, Hill H, Love A, et al. Advances in application of climate prediction in agriculture. Agric Syst. 2001;70(2):515–553. https://doi.org/10.1016/S0308-521X(01)00058-0
7. Byerlee D, Jayne TS, Myers RJ. Managing food price risks and instability in a liberalizing market environment: Overview and policy options. Food Policy. 2006;31(4):275–287. https://doi.org/10.1016/j.foodpol.2006.02.002
8. Becker-Reshef I, Justice C, Sullivan M, Vermote E, Tucker C, Anyamba A, et al. Monitoring global croplands with coarse resolution earth observations: The Global Agriculture Monitoring (GLAM) project. Remote Sens. 2010;2(6):1589–1609. https://doi.org/10.3390/rs2061589
9. Bastiaanssen WG, Ali S. A new crop yield forecasting model based on satellite measurements applied across the Indus Basin, Pakistan. Agric Ecosyst Environ. 2003;94(3):321–340. https://doi.org/10.1016/S0167-8809(02)00034-8
10. Aparicio N, Villegas D, Araus JL, Casadesus J, Royo C. Relationship between growth traits and spectral vegetation indices in durum wheat. Crop Sci. 2002;42(5):1547–1555. https://doi.org/10.2135/cropsci2002.1547
11. Moriondo M, Maselli F, Bindi M. A simple model of regional wheat yield based on NDVI data. Eur J Agron. 2007;26(3):266–274. https://doi.org/10.1016/j.eja.2006.10.007
12. Lopresti MF, Di Bella CM, Degioanni AJ. Relationship between MODIS-NDVI data and wheat yield: A case study in Northern Buenos Aires province, Argentina. Inf Process Agric. 2015;2(2):73–84. https://doi.org/10.1016/j.inpa.2015.06.001
13. Giunta F, Motzo R, Deidda M. Effect of drought on yield and yield components of durum wheat and triticale in a Mediterranean environment. Field Crops Res. 1993;33(4):399–409. https://doi.org/10.1016/0378-4290(93)90161-F
14. Royo C, Voltas J, Romagosa I. Remobilization of pre-anthesis assimilates to the grain for grain only and dual-purpose (forage and grain) triticale. Agron J. 1999;91(2):312–316. https://doi.org/10.2134/agronj1999.00021962009100020021x
15. Wheeler TR, Hong TD, Ellis RH, Batts GR, Morison JI, Hadley P. The duration and rate of grain growth, and harvest index, of wheat (*Triticum aestivum* L.) in response to temperature and CO_2. J Exp Bot. 1996;47(5):623–630. https://doi.org/10.1093/jxb/47.5.623
16. Carlson TN, Ripley DA. On the relation between NDVI, fractional vegetation cover, and leaf area index. Remote Sens Env. 1997;62(3):241–252. https://doi.org/10.1016/S0034-4257(97)00104-1
17. Singh RA, Semwal DP, Rai A, Chhikara RS. Small area estimation of crop yield using remote sensing satellite data. Int J Remote Sens. 2002;23(1):49–56. https://doi.org/10.1080/01431160010014756
18. Reed BC, Brown JF, VanderZee D, Loveland TR, Merchant JW, Ohlen DO. Measuring phenological variability from satellite imagery. J Veg Sci. 1994;5(5):703–714. https://doi.org/10.2307/3235884
19. Tucker CJ. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sens Env. 1979;8(2):127–150. https://doi.org/10.1016/0034-4257(79)90013-0
20. Asner GP, Scurlock JM, Hicke JA. Global synthesis of leaf area index observations: Implications for ecological and remote sensing studies. Glob Ecol Biogeogr. 2003;12(3):191–205. https://doi.org/10.1046/j.1466-822X.2003.00026.x
21. Chen PY, Fedosejevs G, Tiscareno-Lopez M, Arnold JG. Assessment of MODIS-EVI, MODIS-NDVI and VEGETATION-NDVI composite data using agricultural measurements: An example at corn fields in western Mexico. J Env Monit Ass. 2006;119(1–3):69–82. https://doi.org/10.1007/s10661-005-9006-7
22. Yang C, Everitt JH, Bradford JM. Comparison of QuickBird satellite imagery and airborne imagery for mapping grain sorghum yield patterns. Precis Agric. 2006;7(1):33–44. https://doi.org/10.1007/s11119-005-6788-0
23. Nuarsa IW, Nishio F, Hongo C. Rice yield estimation using Landsat ETM+ data and field observation. J Agric Sci. 2011;4(3):45–56. https://doi.org/10.5539/jas.v4n3p45
24. Mutanga S, Van Schoor C, Olorunju PL, Gonah T, Ramoelo A. Determining the best optimum time for predicting sugarcane yield using hyper-temporal satellite imagery. Adv Remote Sens. 2013;2(03):269–275. https://doi.org/10.4236/ars.2013.23029
25. Unganai LS, Kogan FN. Drought monitoring and corn yield estimation in Southern Africa from AVHRR data. Remote Sens Env. 1998;63(3):219–232. https://doi.org/10.1016/S0034-4257(97)00132-6
26. Frost C, Thiebaut N, Newby T. Evaluating Terra MODIS satellite sensor data products for maize yield estimation in South Africa. S Afr J Geom. 2013;2(2):106–119.
27. Hensley M, Le Roux P, Du Preez C, Van Huyssteen C, Kotze E, Van Rensburg L. Soils: The Free State's agricultural base. S Afr Geogr J. 2006;88(1):11–21. https://doi.org/10.1080/03736245.2006.9713842
28. Agricultural Research Council. Guideline for the production of small grains in the summer rainfall areas. Bloemfontein: University of the Free State; 2014. Available from: http://www.arc.agric.za/arc-sgi/
29. Dube E, Mare-Patose R, Kilian W, Barnard A, Tsilo TJ. Identifying high-yielding dryland wheat cultivars for the summer rainfall area of South Africa. S Afr J Plant Soil. 2016;33(1):77–81. https://doi.org/10.1080/02571862.2015.1061712

30. Vermote EF, El Saleous NZ, Justice CO. Atmospheric correction of MODIS data in the visible to middle infrared: First results. Remote Sens Env. 2002;83(1):97–111. https://doi.org/10.1016/S0034-4257(02)00089-5

31. Vermote EF, El Saleous N, Justice CO, Kaufman YJ, Privette JL, Remer L, et al. Atmospheric correction of visible to middle-infrared EOS-MODIS data over land surfaces: Background, operational algorithm and validation. J Geophys Res Atmos. 1997;102(D14):17131–17141. https://doi.org/10.1029/97JD00201

32. Dwyer J, Schmidt G. The MODIS reprojection tool. In: Qu JJ, Gao W, Kafatos M, Murphy RE, Salomonson VV, editors. Earth science satellite remote sensing. Berlin: Springer; 2006. p. 162–177. https://doi.org/10.1007/978-3-540-37294-3_9

33. Savitzky A, Golay MJ. Smoothing and differentiation of data by simplified least squares procedures. Anal Chem. 1964;36(8):1627–1639. https://doi.org/10.1021/ac60214a047

34. Kim SR, Prasad AK, El-Askary H, Lee WK, Kwak DA, Lee SH, et al. Application of the Savitzky–Golay filter to land cover classification using temporal MODIS vegetation indices. Photogramm Eng Remote Sensing. 2014;80(7):675–685. https://doi.org/10.14358/PERS.80.7.675

35. Mkhabela MS, Mkhabela MS, Mashinini NN. Early maize yield forecasting in the four agro-ecological regions of Swaziland using NDVI data derived from NOAA's-AVHRR. Agric Forest Meteorol. 2005;129(1):1–9. https://doi.org/10.1016/j.agrformet.2004.12.006

36. Franch B, Vermote EF, Becker-Reshef I, Claverie M, Huang J, Zhang J, et al. Improving the timeliness of winter wheat production forecast in the United States of America, Ukraine and China using MODIS data and NCAR growing degree day information. Remote Sens Env. 2015;161:131–148. https://doi.org/10.1016/j.rse.2015.02.014

37. Larsen WA, McCleary SJ. The use of partial residual plots in regression analysis. Technometrics. 1972;14(3):781–790. https://doi.org/10.1080/00401706.1972.10488966

38. Ben MG, Yohai VJ. Quantile–quantile plot for deviance residuals in the generalized linear model. J Comput Graph Stat. 2012;13(1):36–47. https://doi.org/10.1198/1061860042949_a

39. Zadoks JC, Chang TT, Konzak CF. A decimal code for the growth stages of cereals. Weed Res. 1974;14(6):415–421. https://doi.org/10.1111/j.1365-3180.1974.tb01084.x

40. Hayes MJ, Decker WL. Using satellite and real-time weather data to predict maize production. Int J Biometeorol. 1998;42(1):10–15. https://doi.org/10.1007/s004840050077

41. Ma BL, Dwyer LM, Costa C, Cober ER, Morrison MJ. Early prediction of soybean yield from canopy reflectance measurements. Agron J. 2001;93(6):1227–1234. https://doi.org/10.2134/agronj2001.1227

42. Benedetti R, Rossini P. On the use of NDVI profiles as a tool for agricultural statistics: The case study of wheat yield estimate and forecast in Emilia Romagna. Remote Sens Env. 1993;45(3):311–326. https://doi.org/10.1016/0034-4257(93)90113-C

43. Mkhabela MS, Bullock P, Raj S, Wang S, Yang Y. Crop yield forecasting on the Canadian Prairies using MODIS NDVI data. Agric Forest Meteorol. 2011;151(3):385–393. https://doi.org/10.1016/j.agrformet.2010.11.012

44. Oyoshi K, Tomiyama N, Okumura T, Sobue S. Mapping rice-planted areas using time-series synthetic aperture radar data for the Asia-RiCE activity. Paddy Water Environ. 2015;14(4):463–472. https://doi.org/10.1007/s10333-015-0515-x

The rise and fall of dissolved phosphate in South African rivers

AUTHOR:
Neil J. Griffin[1]

AFFILIATION:
[1]Unilever Centre for Environmental Water Quality, Institute for Water Research, Rhodes University, Grahamstown, South Africa

CORRESPONDENCE TO:
Neil Griffin

EMAIL:
n.griffin@ru.ac.za

DATES:
Received: 20 Jan. 2017
Revised: 04 May 2017
Accepted: 24 June 2017

KEYWORDS:
eutrophication; nutrients; fresh water; pollution

HOW TO CITE:
Griffin NJ. The rise and fall of dissolved phosphate in South African rivers. S Afr J Sci. 2017;113(11/12), Art. #2017-0020, 7 pages. http://dx.doi.org/10.17159/sajs.2017/20170020

ARTICLE INCLUDES:
× Supplementary material
× Data set

FUNDING:
None

Eutrophication of water resources following nutrient loading is a global threat to water quality, and has been found to be one of the major threats to water quality in South Africa. Eutrophication is large-scale autotroph growth following nutrient enrichment and has several consequences, including loss of biodiversity, oxygen depletion, taste/odour generation and algal toxin production. Phosphate enrichment is often (but not always) the cause of freshwater eutrophication, and limitation of phosphate is commonly used as a means of controlling eutrophication. This study reports on a survey of trends in nutrient levels in South African freshwater resources. The research reported on here shows a significant decrease in dissolved phosphate levels in recent years, following a long period during which phosphate levels had been increasing with time. While changes in inorganic nitrogen were found, these changes did not match those in phosphate levels. Several potential causes of these changes were assessed, and it is concluded that no one cause can explain the changes observed. While the decrease in freshwater phosphate levels bodes well for water quality management, internal phosphorus cycling and other mechanisms are likely to mask the short-term impact of phosphate decreases.

Significance:

- Eutrophication, caused by nutrient loading, is a threat to water quality.
- Phosphate levels in South African fresh water have risen steadily with time, but recently have shown a sharp decrease, which cannot easily be attributed to one phosphate source.
- The decrease has promise for eutrophication management but changes may not occur in the short term.

Introduction

A number of studies have commented on global trends of increasing nutrient loads in fresh water that drive eutrophication of freshwater and coastal ecosystems.[1,2] Eutrophication, recognised as a major threat to surface waters globally, is excessive autotroph growth consequent to nutrient enrichment.[1] Eutrophication has a number of side effects with undesirable ecological and economic consequences that include increased autotroph biomass, species compositional shifts, reductions in biodiversity, potential production of algal toxins, oxygen depletion, and taste and odour production.

Nutrient enrichment effectively removes control over growth and production by limiting nutrients. Although the predominant paradigm suggests that freshwater systems are phosphorus-limited[3,4], both nitrogen and phosphorus enrichment may increase growth in freshwater systems, and concomitant enrichment of nitrogen and phosphorus may have synergistic effects leading to the greatest increase in autotroph production[5,6]. Control of nitrogen in water resources is confounded by the potential of changes to available nitrogen forms and levels as a result of nitrogen cycling, among other factors. Although management of phosphorus levels may be modified by legacy phosphate and other mechanisms[7,8], phosphate limitation is a common approach to managing eutrophication in fresh water. Nutrient limitation is a simple and obvious approach to eutrophication management, and has in some cases been successful.[5]

Nutrient sources that contribute to anthropogenic nutrient loading are many and broadly relate to waste disposal practices and land use and management practices. Globally, the major anthropogenic causes of eutrophication are sewage effluent disposal and agricultural land use management practices.[5,9] Beyond anthropogenic causes, several geochemical and ecological factors affect the quantity and type of nutrients that may be available. As a result of anthropogenic changes, the global nitrogen and phosphorus cycles have increased by factors of about 1 and 4, respectively.[10]

Increasing eutrophication in South African freshwater ecosystems is no exception to the global trend, and a number of publications have expressed concern at the ongoing impact of eutrophication on local water resources.[9,11,12-18] While nutrient levels in many catchments may be low at times, trends show an increase with time in levels of nutrients in most catchments, most particularly in the case of phosphate.[9,19] Increased nutrient levels in South African rivers have been accompanied by increases in eutrophication of reservoirs in the country to the extent that, depending on estimates, half or more of the larger reservoirs in the country are eutrophic or hypertrophic.[12,13,17,18] This level of eutrophication has severe ecological and economic consequences, and concerns have been raised regarding the capacity in South Africa for managing the situation.[18]

Despite a chemical water quality monitoring programme that monitors nutrients at hundreds of points across South Africa, there is relatively little published information on trends in nutrient levels with time. De Villiers and Thiart[9] looked at trends in nitrate/nitrite and phosphate at points low in major catchments over the full data record. Trends in nitrate/nitrite varied from catchment to catchment, and all the statistically significant trends detected were towards decreasing levels of this nutrient. On the other hand, in all but 1 catchment of the 25 assessed, levels of phosphate were found to be increasing – in many cases at a statistically significant rate. This greater

increase in phosphate levels is consistent with the change in nitrogen and phosphorus cycles reported by Falkowski et al.[10]

Since the promulgation of the 1 mg P/L standard for discharged effluent, most of the eutrophication management in South Africa has focused on phosphorus concentration management.[12,18,20,21] Phosphate loading in South Africa has been associated with discharge of wastewater treatment works effluent[19,22,23], urban run-off[24,25], the use of phosphate-containing detergents[12,18,26,27], mining[28,29] and agricultural activity[30,31].

Phosphorus loads from urban areas entering freshwater resources have particularly received considerable attention of late[18] because primary and secondary treatment of wastewater removes very little phosphate[32], and because South African wastewater treatment works do not focus on phosphorus removal[18]. In addition, a significant portion of the phosphorus in domestic wastewater derives from phosphate-containing laundry detergents[26,27] which in many other countries have been banned or limited.

A recent analysis of temporal trends in water quality in two catchments noted an unexpected and sharp decrease in phosphate levels in river water since 2009.[29] This decrease was noted at most sites in the study and largely seemed unrelated to catchment, geology, land use or relative position of the site along the length of the river. The decrease followed many years of increasing phosphate levels. The research described here aims to test whether this observation can be generalised to South Africa as a whole. This paper describes an assessment of historical and recent trends in phosphate and inorganic nitrogen across the country, with consideration of potential drivers.

Methods

Data on major nutrients in rivers and canals collected at the Department of Water and Sanitation's (DWS) 333 priority monitoring points were sourced from the Department's WMS database. Received data were then compiled into a single table (332 084 records from 1952 to 2015, sampling frequency varies but often fortnightly) and processed so that the data format requirements of the various statistical methods were met. Monitoring points without reasonably complete data records (taken as having records from at least 9 months of each year) were omitted from the analysis.

Left-censored data occur where the result is below the detection limit of the analytical approach, and are relatively common in water quality data.[33] Phosphate levels in WMS data contain significant levels of left-censored data, and so appropriate statistical methods were required. Two approaches were selected for the analysis. In the first approach, pairwise tests were used to compare selected years using a paired Prentice–Wilcoxon test as recommended for paired censored data.[34] In the second approach, quantiles in left-censored data were estimated for all years between the first and last year using an implementation of regression on order statistics as described by Helsel[33]. In the latter approach, quantiles were estimated by pooling all data nationally, with no consideration of particular catchment or location.

Assessment of data from Griffin et al.[29] showed a general increase in phosphate levels from the start of the data record until around 2009, whereafter levels dropped sharply. Because the start of the data record at the 333 priority monitoring points varied, the start year for testing was selected as 1985, at which point nearly all points had routine monitoring in place. The mid-year for testing was selected as 2008, after levels had increased to near their peak. The last year for testing was selected as 2013, as this year was well after the decrease noted by Griffin et al.[29], and the latest year in the data set for which data records were complete.

Data on nitrate/nitrite, ammonium and phosphate were selected for analysis as these compounds had the best data record, making them suitable for analysis. Ideally, data on chlorophyll-*a* and total phosphorus (which includes biologically incorporated phosphorus) would be included in a full analysis that addresses nutrient loading, and by implication eutrophication, but the data record on these variables was more sparse after pre-processing. Almost no chlorophyll-*a* records survived pre-processing, and the full chlorophyll-*a* data set without pre-processing was used instead.

Data on total inorganic nitrogen were compiled from nitrate/nitrite and ammonium data. After conversion from mass concentration to molar concentration, data on the molar nitrogen-to-phosphorus ratio (N:P ratio) were derived using total inorganic nitrogen and phosphate data. In both these calculations, propagation of nulls owing to missing data in the contributing data sets increased the amount of null data in the derived indices.

For pairwise testing of the years 1985, 2008 and 2013, pairwise tests were undertaken such that data from the same monitoring point and month were compared across years. Where more than one sample was taken per month, only the first sample in the month was selected for analysis. Monitoring points selected for analysis were those for which data were present for at least 9 months of every year. The locations of the selected monitoring points are presented in Figure 1.

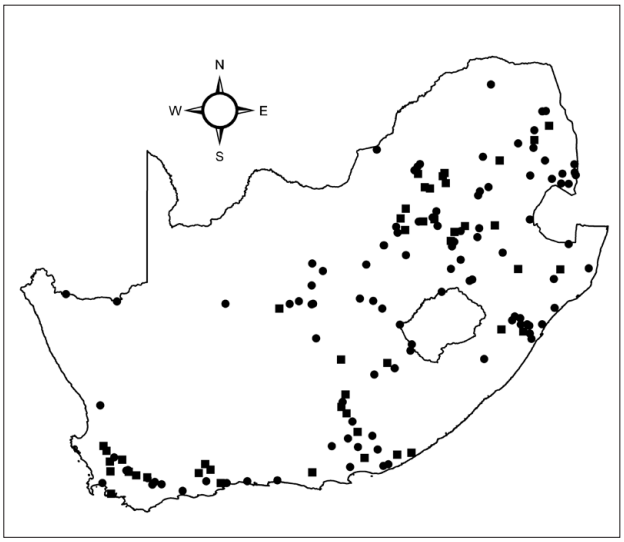


Figure 1: Location in South Africa of monitoring points used in the current analysis. Circles indicate points from which data were used to derive estimates of national nutrient levels. Squares are the subset of these points that had data for 9 months or more of the year in 1985, 2008 and 2013.

All data manipulation and analysis was done using PostgreSQL and R 3.0.2.[35] Regression on Order Statistics made use of routines provided in the package NADA[36], and Prentice–Wilcoxon tests made use of the package smwrQW (Lorenz, in preparation). Other packages used in loading and manipulation of data and plotting of figures were RODBC[37], xlsx[38], plyr[39], ggplot2[40], scales[41] and grid[35].

Results

Estimates of annual median South African phosphate levels are presented in Figure 2. It can be seen that dissolved phosphate levels in South Africa showed an increasing trend that peaked around 2007, whereafter levels decreased to a level comparable to levels encountered during the 1980s. This movement equates to a median increase from 0.016 mg P/L in 1985 to a peak of 0.036 mg P/L in 2007 – more than doubling the amount of phosphate available for uptake by freshwater biota. Thereafter, levels fell swiftly to a median of 0.007 mg P/L in 2012. Pairwise comparisons support this trend, and confirm that at sites across the country, phosphate levels increased from 1985 to 2008 ($p<0.001$), then decreased until 2013 ($p<0.001$). Following the DWS's generic classification, phosphate levels would have been classed as tolerable in 1985, worse than tolerable in 2007, and acceptable in 2013.[15] The 75th percentile is, however, not tolerable, indicating that a large number of sites still have high phosphate levels.

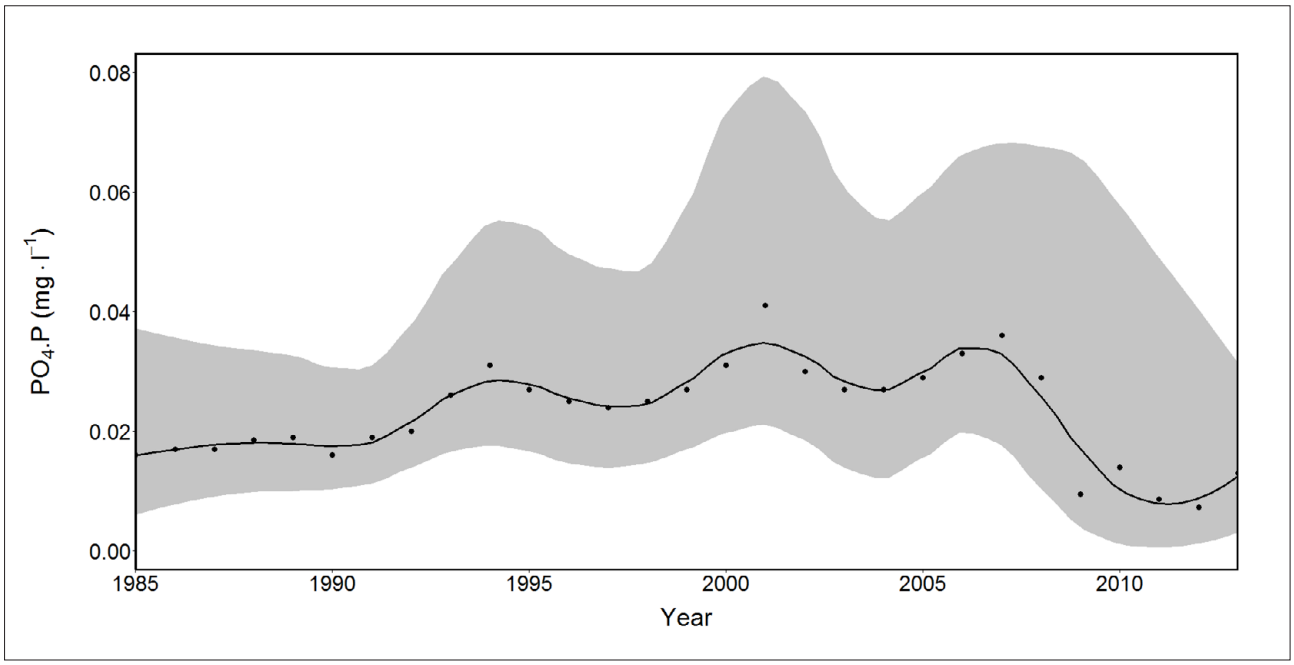


Figure 2: Annual median phosphate levels in South African fresh water plotted against time. The points show median dissolved phosphate (as mg P/L) per year; the line is a loess smoother; and the grey ribbon spans the 25th to 75th percentiles.

Levels of total phosphorus were also assessed to determine whether the trend that was apparent in dissolved phosphate was also reflected in total phosphorus levels. Total phosphorus levels are generally considered a more reliable indicator of incipient eutrophication than dissolved phosphate.[12,18] The data set for total phosphorus was much smaller than that for dissolved phosphate. There was no overt trend with time in total phosphorus, and certainly not one that matches the trends in dissolved phosphate (data not shown). Median estimates of annual total phosphorus ranged from 0.05 mg/L to 0.11 mg/L. This range would be classed as holding significant potential for algal and plant productivity.[12]

A proxy or indicator of potential total phosphorus may be found in assessment of chlorophyll-*a* levels in conjunction with dissolved phosphate, which follows the premise that algal uptake of phosphate removes it from the water column, and algal phosphorus contributes to total phosphorus. Chlorophyll-*a* data from DWS priority monitoring points were not collected evenly across the country, and the data in the data set were collected from 1990 onwards and overwhelmingly from the Orange River. Because of limited data, all data were assessed without following the data filtration processes undertaken for nutrient data. Despite this omission, these chlorophyll-*a* levels were examined to see how they relate to phosphate levels as seen in Figure 2. Chlorophyll-*a* levels showed a decrease after 2007 that paralleled that observed in dissolved phosphate levels (data not shown). Median levels from 2000 (the first year when more than 10 records were collected) to 2007 ranged around 10 μg/L, while after 2009 median levels ranged between 1 μg/L and 4 μg/L. This observation suggests that a large body of phosphorus is not locked up in phytoplankton, and the decrease in chlorophyll-*a* levels indicates a smaller phytoplankton body in recent years. The most parsimonious explanation for this finding is that the decrease in chlorophyll-*a* levels is a function of decreased available phosphate in the resource; however, growth was not phosphate limited across most of the timeframes assessed. It also indicates that a significant amount of phosphorus was held in the phytoplankton body prior to 2007, which suggests in turn that total phosphorus may have been higher before 2007. It should be noted that these observations reflect the state of the Orange River in particular, and may not apply across the entire country. It should also be noted that chlorophyll-*a* data were occasionally censored, but the quantity of data available precluded use of analytic techniques for censored data as undertaken with nutrient data.

Although trends in phosphate levels in South African fresh water are clear, they are not matched by trends in inorganic nitrogen (Figure 3). Levels of nitrogen vary with time, and the only generally directed trend is a slight decrease in median inorganic nitrogen since the start of the data set. The only exception to this trend is a brief period in the late 2000s when elevated levels were encountered. Median inorganic nitrogen levels ranged between around 0.02 mg N/L during 2010 and 0.20 mg N/L during the late 1980s; therefore inorganic nitrogen levels in South African rivers have varied by up to a factor of 10 since 1985. Pairwise comparisons show a slight increase in inorganic nitrogen from 1985 to 2008 ($p=0.015$), and a drop thereafter until 2013 ($p<0.001$). The year 2008 was selected based on the preliminary assessment of phosphate levels. Using data from 2008 is not desirable for inorganic nitrogen levels, as these are drawn from a brief peak in inorganic nitrogen levels that is not present for most of the data record. Nevertheless, for compatibility with the analysis of phosphate levels, it was retained.

These figures may be somewhat low as inspection of annual nitrate/nitrite nitrogen medians reveals these to be slightly higher at times (data not shown). The reason for this discrepancy seems likely to be the smaller sample size used in estimating annual inorganic nitrogen medians, as inorganic nitrogen levels could only be validly calculated when both nitrate/nitrite nitrogen and ammonium nitrogen were sampled and tested for, and detected. Nitrate/nitrite levels indicate that the majority of nitrogen in the inorganic nitrogen pool was in the form of nitrate and nitrite. The upper level for the ideal class used in the DWS generic classification (for nitrate/nitrite nitrogen) is 6 mg N/L.[15] The median levels of inorganic nitrogen reported on here are well below that level and so would be classed as ideal.

The N:P ratio is a commonly used ecological stoichiometric measure used to assess relative amounts of available nutrients to ascertain which might be limiting for growth.[42,43] A molar N:P ratio of 16:1, known as the Redfield ratio, indicates that neither nitrogen nor phosphorus alone is limiting. Figure 4 shows annual medians of the N:P ratio in South Africa. As can be seen, annual median N:P ratios only approximate the Redfield ratio in the late 1980s, with a maximum annual median of 20:1 from 1987, indicating that algal and plant growth was limited by phosphate in that year only. The remaining years returned ratios below 16:1, indicating that for the majority of recent years algal growth was nitrogen-limited, and that phosphate was present in excess.

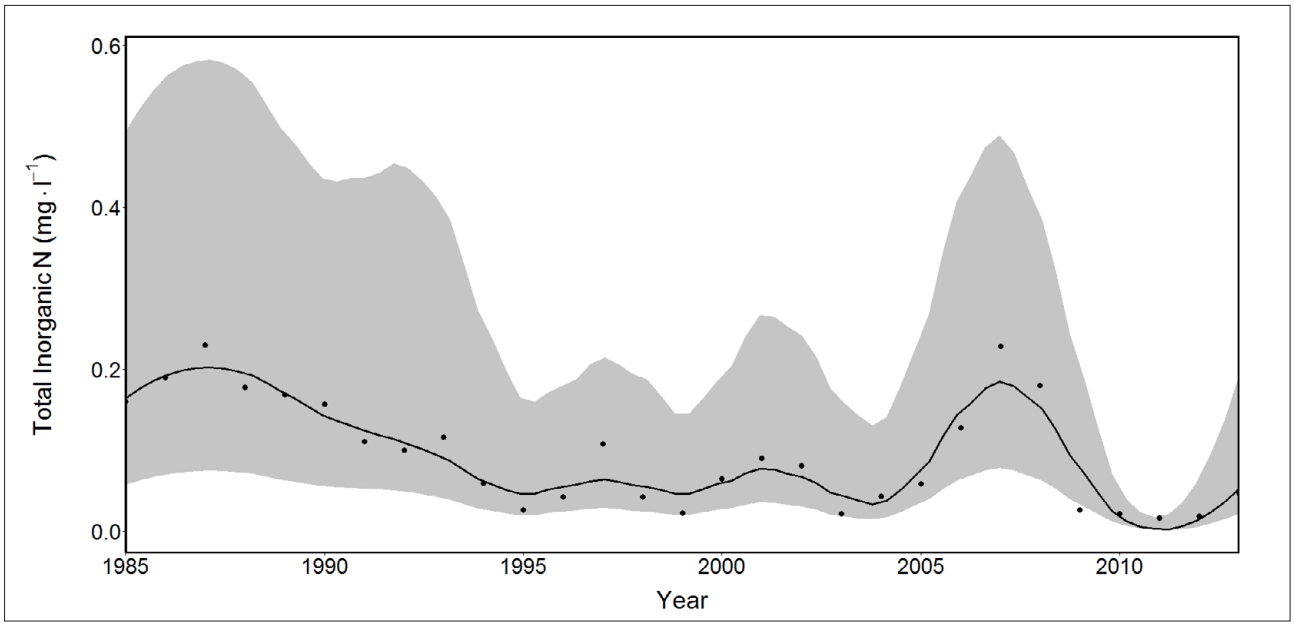


Figure 3: Annual median total inorganic nitrogen levels in South African fresh water plotted against time. The points show median dissolved total inorganic nitrogen (as mg N/L) per year; the line is a loess smoother; and the grey ribbon spans the 25th to 75th percentiles.

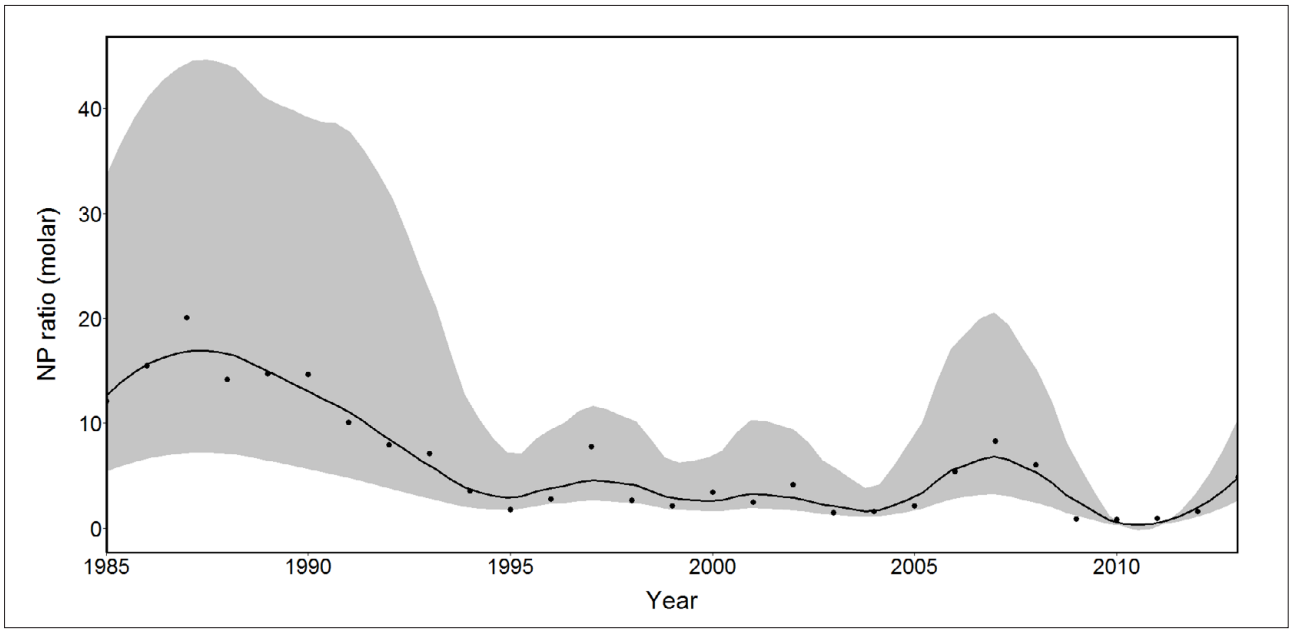


Figure 4: Annual median molar nitrogen-to-phosphorus (NP) ratio in South African fresh water plotted against time. The points show the median molar nitrogen-to-phosphorus ratio per year; the line is a loess smoother; and the grey ribbon spans the 25th to 75th percentiles.

The recent decrease in phosphate levels seen in Figure 2 is linked to a largely unchanging N:P ratio, and, with a ratio of 4.9:1 in 2013, growth remains nitrogen-limited in 2013 – the last year of the data set. This is similar to ratios reported from sewage, zooplankton excreta, and pastureland run-off by Downing and McCauley[42], and below ratios reported from urban stormwater and cattle manure seepage.

Rainfall levels have been found to modify water chemistry by two major mechanisms. The first is increased dilution capacity of water bodies when rainfall is higher and more water is available to dilute dissolved materials, thereby reducing the concentration of dissolved compounds present. The second is the 'first-flush' effect, in which increased surface run-off and baseflow following rainfall may wash compounds into a river, thereby increasing their concentration in the river. In this light, annual rainfall data were assessed to see whether they might explain the observed trends in nutrients. At this point it is worth noting that phosphorus and nitrogen nutrients are involved in biological cycling, which alters the quantities of nutrients available. Rainfall data were sourced from the World Bank and rainfall anomaly data from the South African Weather Service. Annual rainfall data ranged from 320 mm/year to 587 mm/year, with the wettest years being 2000, 2006, 2011, 1988 and 1996, and the driest years being 1992, 2003, 1994, 1990 and 2007. No overall trend is apparent that can simply explain the recent decreases in phosphate that were observed. In addition, no simple correlation between wet or dry years and decreased or increased nutrient levels was apparent. It may be of some significance that increasing phosphate levels from 1990 to 1994 occurred during a period that included three of the driest years in this period (Figure 1); inorganic nitrogen showed no obvious change at this time (Figure 3). However, other short-term increases of phosphate, such as that from 1999 to 2001, occurred when

rainfall was above average. However, the possibility that local variation that more closely links nutrient levels and rainfall is masked in this national assessment cannot be discounted.

Discussion

The data presented here support the conclusion that after a long increase to levels that would be classified as not tolerable, phosphate levels after 2008 dramatically decreased on a national scale. It is also notable that the change in dissolved phosphate was rapid, with the major changes occurring in only 1 year. Because the change happened at a national scale, changes at a more local scale are less likely to be drivers. This raises the question as to what drove this rapid shift, and maintained the recent lowered levels of phosphate.

As noted earlier, the major widespread sources of phosphate in fresh water in South Africa include agricultural effluent and sewage, with a significant phosphate load coming from detergents, especially washing powders. Agricultural effluent could comprise effluent from stocking-intensive operations, nutrient-enriched returns from agricultural operations, and other practices that release nutrients to ground- or surface water. Phosphate is a highly charged and reactive ion that is strongly sorbed by most sediments and which is capable of binding with a number of metal cations, all of which limit its transport in groundwater.[44,45] In contrast, nitrates are more labile but can be removed by denitrification.[46] Nevertheless, phosphates can move through soil, albeit slowly[44-45], and can be mobilised depending on soil and groundwater chemistry[47]. The implication is that while nitrogen can leach from soil where agricultural lands are fertilised or nutrient contaminated, phosphate leaching is likely to be more limited, but can occur, particularly when nutrient application has been prolonged. Phosphate input from agricultural regions can also occur as surface water run-off. This contribution may be significant when considering common thunderstorm rainfall events in South Africa, in which soil that has been fertilised is washed off into water resources, together with any adsorbed phosphorus compounds.

Over the period 2007–2008, the price of phosphate rock increased dramatically.[48] Prices recovered in 2009, although at a higher level than before 2007. The majority of phosphate rock is used in the production of fertiliser, although smaller amounts are used in detergent production, feed additives and chemical production.[49] The increased phosphate rock price was accompanied by an increased fertiliser price, both internationally[48] and locally[50]. Fertiliser consumption (in kilograms per hectare of arable land) in South Africa decreased slightly in 2008, although the magnitude did not approach that of the price increases (data from World Bank[51]). After 2008, fertiliser consumption normalised in 2009, fell further in 2010, and then normalised thereafter. Overall, fertiliser consumption in South Africa has changed little since 2006 and is far below world average consumption. Changes in fertiliser application seem not to be the cause of the decreased riverine phosphate levels.

Much has been written of the negative impact of South African wastewater treatment works on river and reservoir trophic status.[14,15,21-23] Smaller wastewater treatment works, which often have no phosphate removal process and are more subject to failure, pose a greater risk in this regard than do larger operations.[52] The Green Drop certification programme is a benchmarking initiative that commenced in 2008 and aims to improve wastewater management in the country.[53] Results from the scores have shown better service delivery with time, but the improvement has not been dramatic.[53,54] Given the recent start, this programme cannot assess changes that happened before 2008. However, inspection of changes in inorganic nitrogen show that levels after 2008 are about six times lower than those around 2008, and three times lower than those around 2000. In comparison, phosphate levels decreased roughly fourfold from around 2008 to later years. As wastewater treatment plants may produce both nitrogen and phosphate in their effluents[32], it is possible that improvements in wastewater treatment, either in terms of facilities or wastewater treatment works management, may have contributed to the decrease in phosphate observed recently. The relatively unchanged N:P ratio in the years since 2009 supports the likelihood that decreases in both nutrients may have the same cause. However, this finding is not conclusive as other causes may have contributed to the observed changes.

The impacts of wastewater treatment works on phosphate in rivers is a consequence of sewage phosphorus and detergent phosphorus entering the facility. Quayle et al.[27] estimated that detergent phosphorus made up 32% of the total phosphorus found in domestic sewage, and contributed up to 30% of the phosphorus loading of receiving reservoirs.

A major producer of laundry detergent in the country removed builder phosphorus from all its products in 2010[18,27], and a reduction in wastewater effluent phosphorus at the majority of plants assessed was noted a year later[18]. As noted above, this removal would have led to a reduction of phosphate loading in receiving water bodies on a national scale, and may partially explain the observed reduction in phosphate. However, the reduction of phosphate reported here commenced in 2008 or 2009, and removal of phosphorus from laundry detergents in 2010 cannot account for these changes alone, although it is possible that the phosphorus removal process started earlier. It seems likely that a combination of reductions in nutrient loading from wastewater treatment plants and the removal of builder phosphate from detergents has contributed to the decrease in phosphate noted here.

It has long been known that inorganic phosphate in reservoirs is subject to rapid uptake by biota, or adsorption to silt, and transfer to the sediments by sedimentation.[55] On the face of it, it would seem that reduced phosphate loads together with sedimentation would remove phosphate from water bodies and reduced phosphate loads would lead to a rapid reduction in eutrophication. However, internal phosphorus cycling results in the release of phosphate from the sediment, with rates modified by various physicochemical and biotic parameters.[56-59] Sediment phosphorus is found in a wide range of compounds, and bioavailability varies among them.[60] Sediment phosphate loading and release may significantly delay the recovery of eutrophic systems once phosphate loading is limited.[56]

In conclusion, levels of phosphate have recently dropped in South African surface waters – a trend largely matching that seen more recently in inorganic nitrogen. During this change, levels fell from a point at which national guidelines were exceeded to a more acceptable level. Despite the changes, phosphate levels, although tenfold lower, remain relatively higher than inorganic nitrogen levels. These trends were found on a national scale, and local variation may mask the trend in places. Given the links between phosphate levels and eutrophication, decreased phosphate loading bodes well for eutrophication management, although internal phosphorus cycling will mean that eutrophic and hypereutrophic water bodies will not recover in the immediate future. Nevertheless, the recent period with decreased phosphate levels has been relatively short, and no firm conclusions as to long-term phosphate levels can be made as yet.

Acknowledgements

This study was based on observations made during projects funded by the Water Research Commission and USAID. The Centre receives additional funding from Unilever. Data are from the Department of Water and Sanitation's WMS database. The World Bank and the South African Weather Service provided rainfall data; D.C.H. Retief contributed the observation that phosphate levels decreased as the price of phosphate rock increased; S. Vetter reviewed the draft manuscript; and two anonymous reviewers provided valuable input.

References

1. Smith VH, Schindler DW. Eutrophication science: Where do we go from here? Trends Ecol Evol. 2009;24(4):201–207. https://doi.org/10.1016/j.tree.2008.11.009
2. Seitzinger SP, Mayorga E, Bouwman AF, Kroeze C, Beusen AHW, Billen G, et al. Global river nutrient export: A scenario analysis of past and future trends. Global Biogeochem Cy. 2010;24, GB0A08. https://doi.org/10.1029/2009gb003587
3. Schindler DW. Recent advances in the understanding and management of eutrophication. Limnol Oceanogr. 2006;51:356–363. https://doi.org/10.4319/lo.2006.51.1_part_2.0356

4. Smith VH. Responses of estuarine and coastal marine phytoplankton to nitrogen and phosphorus enrichment. Limnol Oceanogr. 2006;51:377–384. https://doi.org/10.4319/lo.2006.51.1_part_2.0377

5. Smith VH, Tilman GD, Nekola JC. Eutrophication: Impacts of excess nutrient inputs on freshwater, marine, and terrestrial ecosystems. Environ Pollut. 1999;100:179–196. https://doi.org/10.1016/S0269-7491(99)00091-3

6. Elser JJ, Bracken MES, Cleland EE, Gruner DS, Harpole WS, Hillebrand H, et al. Global analysis of nitrogen and phosphorus limitation of primary producers in freshwater, marine and terrestrial ecosystems. Ecol Lett. 2007;10:1135–1142. https://doi.org/10.1111/j.1461-0248.2007.01113.x

7. Sharpley S, Jarvie HP, Buda A, May L, Spears B, Kleinman P. Phosphorus legacy: Overcoming the effects of past management practices to mitigate future water quality impairment. J Environ Qual. 2013;42:1308–1326. https://doi.org/10.2134/jeq2013.03.0098

8. Zamparas M, Zacharias I. Restoration of eutrophic freshwater by managing internal nutrient loads. A review. Sci Total Environ. 2014;496:551–562. https://doi.org/10.1016/j.scitotenv.2014.07.076

9. De Villiers S, Thiart C. The nutrient status of South African rivers: Concentrations, trends and fluxes from the 1970s to 2005. S Afr J Sci. 2007;103:343–349.

10. Falkowski P, Scholes RJ, Boyle E, Canadell J, Canfield D, Elser J, et al. The global carbon cycle: A test of our knowledge of earth as a system. Science. 2000;290:291–296. https://doi.org/10.1126/science.290.5490.291

11. Hohls BC, Silberbauer MJ, Kühn AL, Kempster PL, Van Ginkel CE. National water resource quality status report: Inorganic chemical water quality of surface water resources in SA – the big picture. Report no. N/0000/REQ0801. Pretoria: Institute for Water Quality Studies, Department of Water Affairs and Forestry; 2002.

12. Van Ginkel CE. Eutrophication: Present reality and future challenges for South Africa. Water SA. 2011;37(5):693–701. https://doi.org/10.4314/wsa.v37i5.6

13. Van Ginkel CE. Algae, phytoplankton and eutrophication research and management in South Africa: Past, present and future. Afr J Aquat Sci. 2012;37(1):17–25. https://doi.org/10.2989/16085914.2012.665432

14. Council for Scientific and Industrial Research (CSIR). A CSIR perspective on water in South Africa – 2010. CSIR report no. CSIR/NRE/PW/IR/2011/0012/A. Pretoria: CSIR; 2010.

15. South African Department of Water Affairs (DWA). Planning level review of water quality in South Africa. Sub-series no. WQP 2.0. Pretoria : DWA; 2011.

16. South African Department of Water Affairs (DWA). The annual national state of water resources report October 2011 to September 2012. Pretoria: DWA; 2011.

17. Matthews MW. Eutrophication and cyanobacterial blooms in South African inland waters: 10 Years of MERIS observations. Remote Sens Environ. 2014;155:161–177. https://doi.org/10.1016/j.rse.2014.08.010

18. Harding WR. Living with eutrophication in South Africa: A review of realities and challenges. Trans Roy Soc S Afr. 2015;70(2):155–171. https://doi.org/10.1080/0035919X.2015.1014878

19. Dabrowski JM, De Klerk LP. An assessment of the impact of different land use activities on water quality in the upper Olifants River catchment. Water SA. 2013;39(2):231–244. https://doi.org/10.4314/wsa.v39i2.6

20. Rossouw JN. Final report on the development of management-orientated models for eutrophication control. WRC report no. 174/1/90. Pretoria: Water Research Commission; 1990.

21. Harding WR. The determination of annual phosphorus loading limits for South African dams. WRC report no. 1687/1/08. Pretoria: Water Research Commission; 2008.

22. Momba MNB, Osode AN, Sibewu M. The impact of inadequate wastewater treatment on the receiving water bodies – Case study: Buffalo City and Nkokonbe Municipalities of the Eastern Cape Province. Water SA. 2006;32(5):687–692.

23. Oberholster PJ, Botha A-M, Chamier J, De Klerk AR. Longitudinal trends in water chemistry and phytoplankton assemblage downstream of the Riverview WWTP in the Upper Olifants River. Ecohydrol Hydrobiol. 2013;13:41–51. https://doi.org/10.1016/j.ecohyd.2013.03.001

24. Vos AT, Roos JC. Causes and consequences of algal blooms in Loch Logan, an urban impoundment. Water SA. 2005;31(3):385–392.

25. Nyenje PM, Foppen JW, Uhlenbrook S, Kulabako R, Muwanga A. Eutrophication and nutrient release in urban areas of sub-Saharan Africa – A review. Sci Total Environ. 2010;4–8:447–455. https://doi.org/10.1016/j.scitotenv.2009.10.020

26. Pillay M. Detergent phosphorus in South Africa: Impact on eutrophication with specific reference to the Umgeni Catchment [MSc thesis]. Durban: University of Natal; 1994.

27. Quayle LM, Dickens CWS, Graham M, Simpson D, Goliger A, Dickens JK, et al. Investigation of the positive and negative consequences associated with the introduction of zero-phosphate detergents into South Africa. WRC report no. TT 446/10. Pretoria: Water Research Commission; 2010.

28. Ashton PJ, Dabrowski JM. An overview of surface water quality in the Olifants River catchment. WRC report no. KV 293/11. Pretoria: Water Research Commission; 2011.

29. Griffin NJ, Palmer CG, Scherman P-A. Critical analysis of environmental water quality in South Africa: Historic and current trends. WRC report no. 2184/1/14. Pretoria: Water Research Commission; 2014.

30. Du Plessis HM, Van Veelen M. Water quality: salinization and eutrophication time series and trends in South Africa. S Afr J Sci. 1991;87:11–16.

31. De Villiers S. The deteriorating nutrient status of the Berg River, South Africa. Water SA. 2007;33(5):659–664.

32. Carey RO, Migliaccio KW. Contribution of wastewater treatment plant effluents to nutrient dynamics in aquatic systems: A review. Environ Manage. 2009;44:205–217. https://doi.org/10.1007/s00267-009-9309-5

33. Helsel D. Statistics for censored environmental data using Minitab and R. 2nd ed. Hoboken, NJ: John Wiley and Sons; 2012.

34. O'Brien PC, Fleming TR. A paired Prentice-Wilcoxon test for censored paired data. Biometrics. 1987;43(1):169–180. https://doi.org/10.2307/2531957

35. R Core Team. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2013. Available from: http://www.R-project.org

36. Lee L. NADA: Nondetects and data analysis for environmental data. R package version 1.5-6. 2013. Available from: http://CRAN.R-project.org/package=NADA

37. Ripley B, Lapsley ML. RODBC: ODBC Database access. R package version 1.3-5. 2012. Available from: http://CRAN.R-project.org/package=RODBC

38. Dragulescu AA. xlsx: Read, write, format Excel 2007 and Excel 97/2000/XP/2003 files. R package version 0.5.7. 2012. Available from: http://CRAN.R-project.org/package=xlsx

39. Wickham H. The split-apply-combine strategy for data analysis. J Stat Softw. 2011;40(1):1–29. https://doi.org/10.18637/jss.v040.i01

40. Wickham H. ggplot2: Elegant graphics for data analysis. New York: Springer; 2009.

41. Wickham H. scales: Scale functions for graphics. R package version 0.2.3. 2012. Available from: http://CRAN.R-project.org/package=scales

42. Downing JA, McCauley E. The nitrogen:phosphorus relationship in lakes. Limnol Oceanogr. 1992;37(5):936–945. https://doi.org/10.4319/lo.1992.37.5.0936

43. Teubner K, Dokulil MT. Ecological stoichiometry of TN:TP:SRSi in freshwaters: Nutrient ratios and seasonal shifts in phytoplankton assemblages. Arch Hydrobiol. 2002;154(4):625–646. https://doi.org/10.1127/archiv-hydrobiol/154/2002/625

44. Robertson WD, Schiff SL, Ptacek CJ. Review of phosphate mobility and persistence in 10 septic system plumes. Ground Water. 1998;36(6):1000–1010. https://doi.org/10.1111/j.1745-6584.1998.tb02107.x

45. Spiteri C, Slomp CP, Regnier P, Meile C, Van Cappellen P. Modelling the geochemical fate and transport of wastewater-derived phosphorus in contrasting groundwater systems. J Contam Hydrol. 2007;92:87–108. https://doi.org/10.1016/j.jconhyd.2007.01.002

46. Rivett MO, Buss SR, Morgan P, Smith JWN, Bemment CD. Nitrate attenuation in groundwater: A review of biogeochemical controlling processes. Water Res. 2008;42:4215–4232. https://doi.org/10.1016/j.watres.2008.07.020

47. Smolders AJP, Lucassen ECHET, Bobbink T, Roelofs JGM, Lamers LPM. How nitrate leaching from agricultural lands provokes phosphate eutrophication in groundwater fed wetlands: The sulphur bridge. Biogeochemistry. 2010;98:1–7. https://doi.org/10.1007/s10533-009-9387-8

48. Ott H. Fertilizer markets and their interplay with commodity and food prices. Report EUR 25392 EN. Seville: Joint Research Centre, European Commission; 2012.

49. Van Vuuren DP, Bouwman AF, Beusen AHW. Phosphorus demand for the 1970–2100 period: A scenario analysis of resource depletion. Global Environ Chang. 2010;20:428–439. https://doi.org/10.1016/j.gloenvcha.2010.04.004

50. Grain SA. Grain SA fertilizer report. Pretoria: Grain SA; 2011.

51. World Bank. Fertilizer consumption (kilograms per hectare of arable land) [document on the Internet]. c2017 [cited 2017 Jan 18]. Available from: http://data.worldbank.org/indicator/AG.CON.FERT.ZS

52. Gaydon P, McNab N, Mulder G, Pillay I, Sahibdeen M, Thompson P. Evaluation of sewage treatment package plants for rural, peri-urban and community use. WRC report no. 1539/1/06. Pretoria: Water Research Commission; 2007.

53. Brettenny W, Sharp G. Efficiency evaluation of urban and rural municipal water service authorities in South Africa: A data envelopment analysis approach. Water SA. 2016;42(1):11–19. https://doi.org/10.4314/wsa.v42i1.02

54. South African Department of Water and Sanitation. Green drop system [homepage on the Internet]. c2017 [cited 2017 Mar 22]. Available from: https://www.dwa.gov.za/dir_ws/gds/

55. Rigler FH. A tracer study of the phosphorus cycle in lake water. Ecology. 1956;37(3):550–562. https://doi.org/10.2307/1930179

56. Ripl W. Internal phosphorus recycling mechanisms in shallow lakes. Lake Reserv Manage. 1986;2(1):138–142. https://doi.org/10.1080/07438148609354616

57. Jensen HS, Andersen FO. Importance of temperature, nitrate, and pH for phosphate release from aerobic sediments of four shallow, eutrophic lakes. Limnol Oceanogr. 1992;37(3):577–589. https://doi.org/10.4319/lo.1992.37.3.0577

58. Reddy KR, Diaz OA, Scinto LJ, Agami M. Phosphorus dynamics in selected wetlands and streams of the lake Okeechobee Basin. Ecol Eng. 1995;5:183–207. https://doi.org/10.1016/0925-8574(95)00024-0

59. Golterman HR. Phosphate release from anoxic sediments or 'What did Mortimer really write?' Hydrobiologia. 2001;450:99–106. https://doi.org/10.1023/A:1017559903404

60. Reitzel K, Ahlgren J, Gogoll A, Jensen HS, Rydin E. Characterization of phosphate in sequential extracts from lake sediments using ^{31}P nuclear magnetic resonance spectroscopy. Can J Fish Aquat Sci. 2006;63:1686–1699. https://doi.org/10.1139/f06-070

The contribution of copyright-based industries to the South African economy

AUTHORS:
Anastassios Pouris[1]
Roula Inglesi-Lotz[2]

AFFILIATIONS:
[1]Institute for Technological Innovation, University of Pretoria, Pretoria, South Africa
[2]Department of Economics, University of Pretoria, Pretoria, South Africa

CORRESPONDENCE TO:
Roula Inglesi-Lotz

EMAIL:
Roula.Inglesi-Lotz@up.ac.za

DATES:
Received: 20 Sep. 2016
Revised: 27 Mar. 2017
Accepted: 04 July 2017

KEYWORDS:
economic impact; intellectual property rights; multipliers; creative industries; WIPO

HOW TO CITE:
Pouris A, Inglesi-Lotz R. The contribution of copyright-based industries to the South African economy. S Afr J Sci. 2017;113(11/12), Art. #2016-0286, 7 pages. http://dx.doi.org/10.17159/sajs.2017/20160286

ARTICLE INCLUDES:
× Supplementary material
× Data set

FUNDING:
World Intellectual Property Organization

We report the results of an effort to measure the contribution of copyright-based industries to the South African economy. Following the methodology of the World Intellectual Property Organization, we identify the copyright industry's contribution to GDP, employment, imports and exports in South Africa for the period 1970–2009. It was estimated that the sector contributed 4.1% to GDP – more than the contributions of other sectors such as agriculture and food, beverages and tobacco. Because of this quantified importance of the copyright-based industries, we recommend that relevant South African policy authorities and policymakers should monitor and publicise regularly the performance of the copyright-based industries as well as promote programmes for their development and growth.

Significance:

- Copyright-based industries have the potential to play an important role in job creation and economic growth.
- Our findings suggest that policymakers should design and promote the implementation of future policies and strategies related to these sectors.

Introduction

In current modern economies, knowledge is considered to be an influential aspect of economic growth and development as it contributes to the improvement of all the other traditional factors of production: capital, labour, technological progress and entrepreneurship. The creation of new knowledge is highly dependent on the protection of intellectual property.[1] In essence, the creation and investment of creative and innovative works should be promoted and directed properly and effectively by each country's copyright laws.[2] However, as specifically discussed by Gani-Ikilama[3], there are numerous and diverse approaches as to how to define creative industries, and thus the investigation of their impact has not reached consensus[4].

Industries protected by copyright and related rights are expected to have considerable impact on national economies. Studies that have examined copyright through a legal lens are numerous[5,6]; Yao and Peng[6], for example, concluded that the leniency of copyright protection may affect the level of benefit thereof for the various stakeholders. However, measuring quantitatively the relevant contribution of copyright-based industries to an economy or region is only a recent phenomenon. Although several international studies have been conducted with regard to the importance of individual copyright industries[7,8], limited efforts have been made to capture and estimate the economic impact of the copyright-based industries in South Africa. In recent years, the World Intellectual Property Organization (WIPO) and the South African Department of Trade and Industry have supported the collection of data including data on copyright-based activities. At the time of our study, it was only the second study in the African region; the first being a Kenyan study[9]. More recently, similar studies have been completed in Ethiopia and Malawi.[10]

Our objective was to measure and report on the economic role of copyright-based industries in the South African economy through the classification defined and used by WIPO.[11] The results show the importance of the various sectors and thus have the potential to inform policymakers on the design and implementation of future policies and strategies towards improving the country's growth and development.

Similar studies quantifying the economic contribution of copyright-based industries of developed and developing economies indicate the importance of these industries. The total economic contribution of copyright-based industries as a percentage to GDP varies from 2.81% in Bulgaria to 11.70% in the Philippines. Similarly, the indicator ratio of persons employed in the copyright-based sector to the total number of employees in the economy varies from 3.03% in Jamaica to 11.17% in Latvia.[11]

Institutional framework on copyright in South Africa

The *Copyright Act, Act No. 98 of 1978* and its amendments legislated copyright for the first time in South Africa. Section 2 of the Act identifies the following as works that are eligible for copyright: literary works, musical works, artistic works, cinematographic films, sound recordings, broadcasts, programme-carrying signals, published editions, and computer programs.[11]

South Africa's commitment to copyright issues and protection of artistic works is demonstrated by the country being party to the *Berne Convention for the Protection of Literary and Artistic Works* and the *Agreement on Trade-Related Aspects of Intellectual Property Rights* (TRIPS), as well as being a signatory of the *WIPO Copyright Treaty*. From a national policy perspective, the country has supported its commitment since 1998. The first strategy dealing with cultural and other copyright-related industries was published by the then Department of Arts, Culture, Science and Technology (DACST), titled *Cultural Industries Growth Strategy*.[12] DACST valued this study as its main contribution to the country's Growth, Employment and Redistribution (GEAR) strategy. The DACST strategic plan categorised the following industries as 'cultural': music, craft, publishing, and film and television. These industries were selected for their potential competitive advantage internationally as well as their potential for the creation of jobs and opportunities for rural and urban growth and development.

The strategy[12] – as summarised in WIPO[11] – makes key recommendations for the future of these industries, stressing the importance of[11]:

- *designing and implementing a Cultural Industries Development Programme (CIDP), setting up a CIDP regulatory framework*
- *creating a Cultural Industries Development Fund*
- *promoting the industries internationally and setting up an export programme*
- *coordinating copyright legislation to protect the local cultural products*
- *developing human resources and skills appropriate to cultural industries*
- *adopting and coordinating government supply side measures*
- *designing and implementing an awareness campaign focused on audience development*
- *collecting and monitoring statistics*
- *coordinating initiatives in other departments*
- *establishing a Cultural Industries Development Agency*.

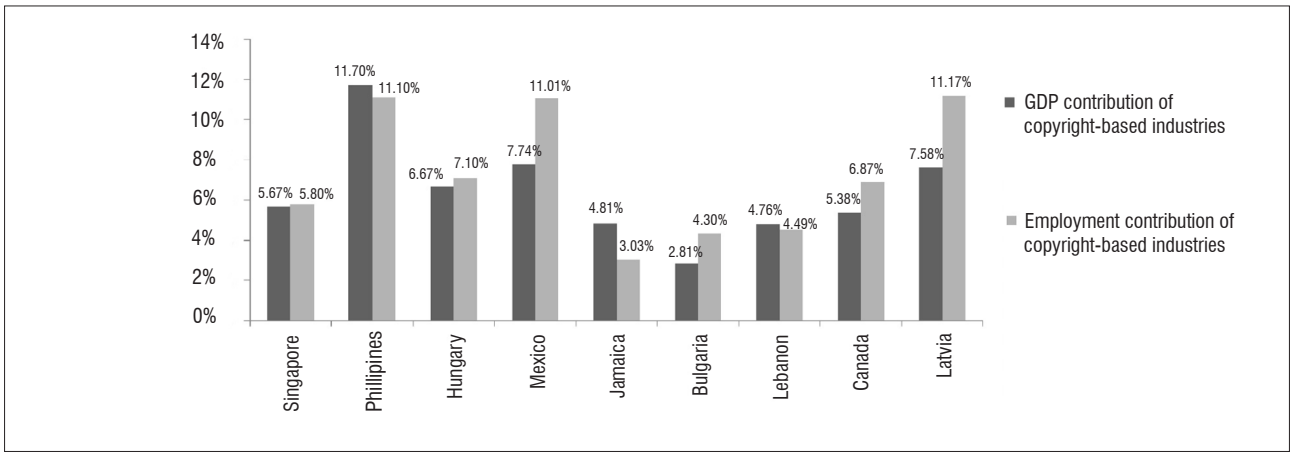
Brief literature review

WIPO[1] notes that 'the economic contribution of the copyright-based industries has exceeded expectations in the last two decades'.

To promote related studies in the policy sphere, WIPO[13] published the *Guide on Surveying the Economic Contribution of the Copyright-based Industries*. The report provides guidance with regard to the methods as well as relevant indicators for such studies. Figure 1 summarises the contribution of copyright-based industries to selected economies internationally (studies commissioned by WIPO). The percentage of economic output ranges from 2.81% for Bulgaria to 11.7% for the Philippines, while that for employment is from 3.03% in Jamaica to 11.17% in Latvia.

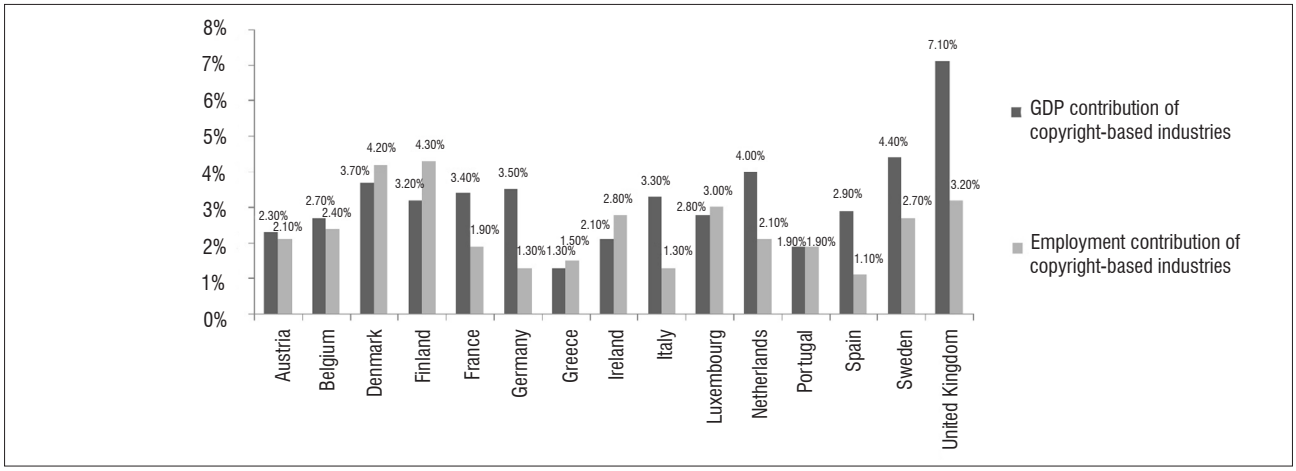
Studies independent of WIPO have confirmed similar results; for example, Siwek[14] found that the US copyright-based industries contributed 7.75% to the country's GDP and 5.9% of the total labour in 2001.

Looking exclusively at European countries, Media Group[15] concluded that copyright-based sectors are imperative for economic production and evolution of the labour force in Europe (Figure 2).



Source: Adapted from WIPO[11]

Figure 1: Contribution of copyright-based industries to GDP and employment in various countries.



Source: Adapted from WIPO[11]

Figure 2: Contribution of copyright-based industries to GDP and employment in countries of the European Union.

Table 1: Classification of copyright-based industries

Core	Interdependent	Partial	Non-dedicated
Printing, publishing and recorded media	Television, radio and communication equipment	Apparel, textiles and footwear	General wholesale and retailing
Film and television industry	Computers and equipment/ photocopiers	Furniture/ jewellery/ musical instruments/ games and toys	Transport, storage and communication
Photography/ software and databases/ advertising	Paper and paper products	Crafts	
Copyright collecting societies		Glass and glass products	

Methodology and data

As per WIPO[13], 'copyright-based industries are those engaged in creation, production and manufacturing, performance, broadcast, communication and exhibition or distribution and sales of works and other protected subject matter'. WIPO[13] explains that 'economic impact can be related to both *core* copyright-based industries and *non-core* industries'.

Relative to their level and type of association to copyright, the main copyright-based industries are categorised into four sub-industries[13]:

- *Core copyright industries: industries wholly engaged in creation, production and manufacturing of performance, broadcast, communication and exhibition or distribution and sales of works and other protected subject matter.*
- *Interdependent copyright industries: industries engaged in production, manufacture and sale of equipment whose function is wholly or primarily to facilitate the creation, production or use of works and other protected subject matter.*
- *Partial copyright industries: industries in which a portion of the activities is related to works and other protected subject matter and may involve creation, production and manufacturing, performance, broadcast, communication and exhibition or distribution and sales.*
- *Non-dedicated support industries: industries in which a portion of the activities is related to facilitating, broadcast, communication, distribution or sales of works or other protected subject matter and whose activities have not been included in the core copyright industries*.

Table 1 presents the South African sub-sectors that were classified as copyright-based industries and includes the specific categories to which they belong.

The basic data sources for the analysis were: Business Monitor International Ltd; Department of Arts, Culture, Science and Technology; Department of Labour; Economist Intelligence Unit; National Organisation for Reproduction Rights in Music in Southern Africa (NORM); Publishers Association; Quantec databases; SA Recording Rights Association Ltd (SARRAL); South African Book Development Council (SABDC); South African Reserve Bank; South African Revenue Service (SARS); Southern African Music Rights Organisation (SAMRO) and Statistics South Africa.[11]

To evaluate and quantify the copyright content for various industries, we used the copyright factor. This factor is defined as the percentage ratio of copyright activities in a given industry. Various methodologies for the estimation of this factor are available in the literature. We used the method of the Singaporean case study[1] (see Table 2) because Singapore is also a newly industrialised economy that depends on its trade and other copyright-related industries. Also, the two countries have similar copyright legislation, intellectual property rights laws and piracy rates.[16,17]

All core copyright-based industries, as shown in Table 1, have a copyright factor of 1 (100%); all the non-dedicated support industries have a copyright factor of 0.057. For the interdependent industries, the copyright factor varies from 0.25 (paper and paper products) to 0.35 (television, radio and communication as well as computers and equipment), while for the partial copyright-based industries, the variation is much higher: from 0.420 (crafts) to 0.004 (apparel, textiles and footwear).

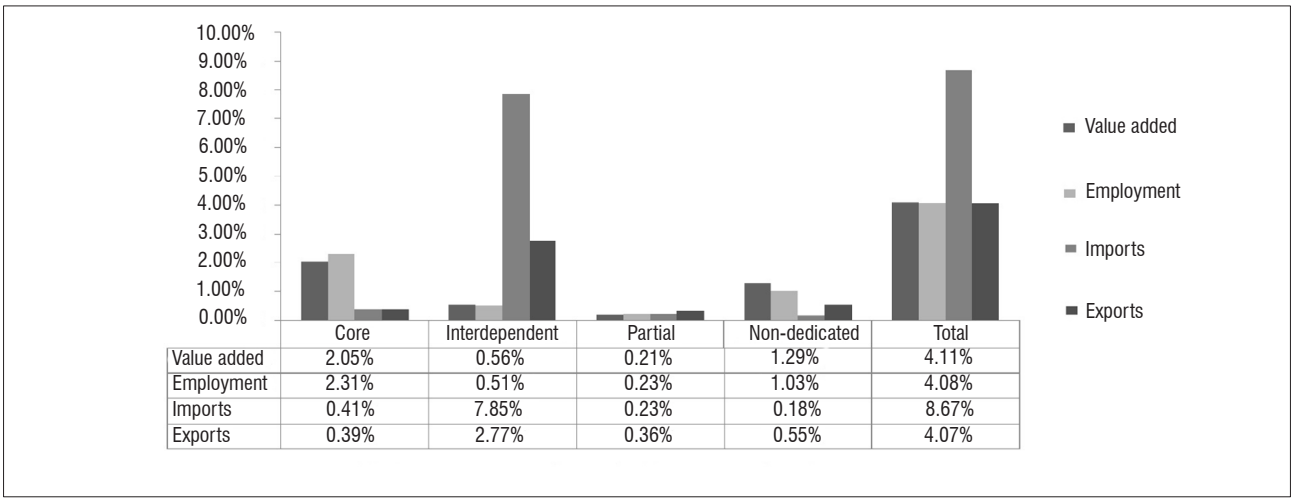
Empirical results

The first part of the quantitative analysis is a presentation of a share of the various indicators to the total economy. Figure 3 quantifies the economic importance of the copyright-based industries in 2009. The total copyright-based industries sector contributes as much as 4.1% to the total South African economy. Among the sub-sectors of copyright-based industries, the core industries contributed 2.05%, the non-dedicated industries 1.29%, the interdependent industries 0.56% and the partial industries 0.21%. The total copyright-based industries employed 4.08% of the country's workforce. As far as exports are concerned, the share of the copyright-based industries was at a similar level (4.07%), and for imports, the total copyright-based industries' contribution was approximately double that of the other indicators (7.85%) – a finding that is not that surprising, especially for countries that have a low high-tech manufacturing base.

The contribution of copyright-based industries in total to the economy's value added showed a constant trend at about 4% for the period 1970–2008 (Figure 4); the same trend was observed in the imports' contribution of the copyright-based industries. In contrast, the industries' contribution to the country's exports and employment showed an increasing trend for the period 1970–2009. The difference in contributions between imports and exports could be attributed to the higher local share of production of interdependent copyright-based industries as a result of the development of advanced technologies in the digital age.

To disaggregate these trends further, Table 2 presents the growth of the shares of each of the various main categories of copyright-based industries to the total economies in terms of value added, employment, imports and exports. These figures are presented in year-on-year growth rates of the contribution. For value added, in all the sub-divisions, the observed trend was growth in the last decades, with the exception of the partial copyright-based industries which experienced a decrease of 33% during the 1990s. In terms of employment, the interdependent copyright industries showed a decrease over the last two decades, while the other industries showed a continuous increase in their share. For imports and exports, the growth rates of the shares of all copyright-based industries were significantly higher than those of the rest of the indicators, showing how volatile the trade of these industries are in South Africa. In addition, the trends followed the overall South African trade, specifically with respect to increases in the 1990s after the end of sanctions.

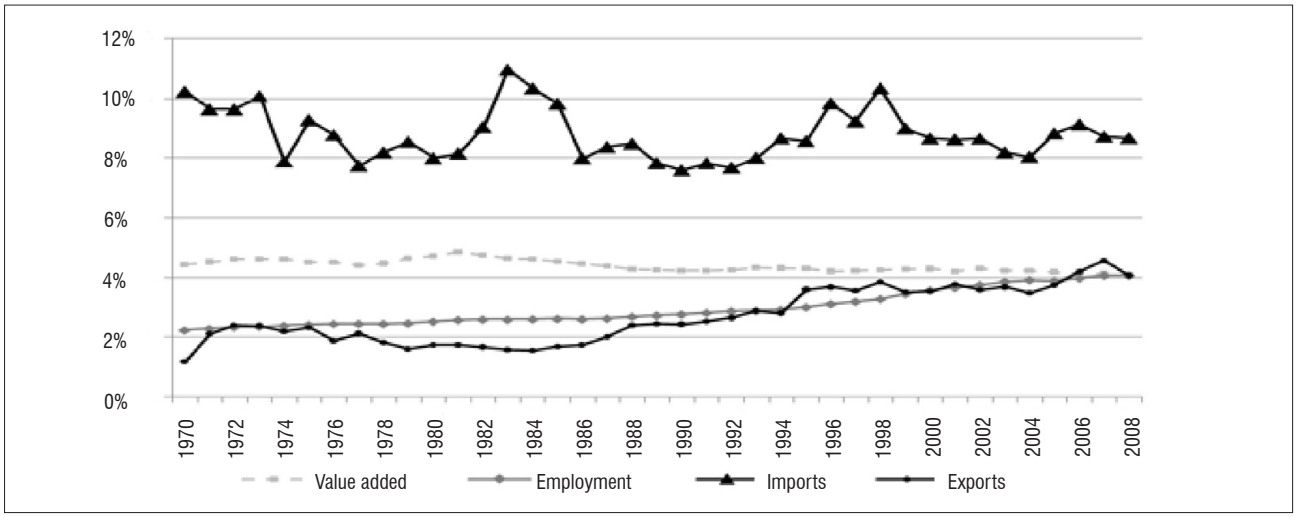
As mentioned, the partial copyright-based industries decreased their share in total value added in the 1990s. Combining this trend with a significant increase in imports (175%) and an increase – albeit slower – in exports (18%) in the same period, one can conclude that the



	Core	Interdependent	Partial	Non-dedicated	Total
Value added	2.05%	0.56%	0.21%	1.29%	4.11%
Employment	2.31%	0.51%	0.23%	1.03%	4.08%
Imports	0.41%	7.85%	0.23%	0.18%	8.67%
Exports	0.39%	2.77%	0.36%	0.55%	4.07%

Source: Adapted from WIPO[11]

Figure 3: Contribution of the different categories of copyright-based industries to the South African economy in 2009.



Source: Adapted from WIPO[11]

Figure 4: Contribution of copyright-based industries to the South African economy (value added, employment, imports and exports) from 1970 to 2008.

industries presented a lack of comparative advantage of their product in comparison with the rest of the world. That might also explain the negative growth in employment during the 2000s.

By comparison, the core copyright industries experienced a decrease (-29%) in imports and a large increase (108%) in exports in the same period, indicating that the sector is relatively competitive with the rest of the world. In addition, the growth rate of the employment share has steadily increased through the decades – almost hand in hand with the value added trends.

Even though the share of value added of the interdependent industries showed an increasing trend through the years, the employment share showed a negative growth during the last two decades of the period under assessment. The reason for this finding might be that these industries became more capital intensive (with less need for human capital). Globalisation, technological progress and know-how transfer could be reasons for that transformation. So why were the other industries not significantly affected by these phenomena? Possibly because the other copyright-based industries are less technology driven.

A basic input-output analysis was used to estimate the direct and indirect effects of the total copyright-based industries to the South African economy. Table 3 shows the production-induced effect of a number of industries on total output. The multiplier illustrates how much a sector should increase its inputs (both internally and other sectors) to produce one extra unit of product to meet an increase in final demand of ZAR1 (first-round effects). The industrial support effect shows how much various industries should increase their purchase to increase their output by that much to meet the first-round requirements.[11] The figures show that the copyright-based industries are not a negligible factor/sector in the South African economy and can actually make a difference.

To put these findings in perspective, the analysis proceeds with a comparison with other main sectors in South Africa's economy (Figure 5). Copyright-based industries contributed, unsurprisingly, less to value added than manufacturing (18.36%) and mining (6.15%), but more than agriculture (2.51%) and food, beverages and tobacco (3.09%) in 2008. The copyright-based industries employed less labour than the mining, manufacturing and agriculture sectors (4.96%, 12.47% and 7.27% respectively) but more than the food, beverages and tobacco sector (1.96%) in the same year. Finally, with respect to the trade potential of the copyright-based industries, these industries contributed more (4.07%) to the country's exports than the agriculture and food, beverages and tobacco sectors (both below 4%), although still not an amount comparable with those of the main exporters of the country: manufacturing and mining. Similar results emerge from the imports comparison.

Table 2: Copyright-based industries' growth of share to the economy of value added, employment, imports and exports from 1970 to 2008

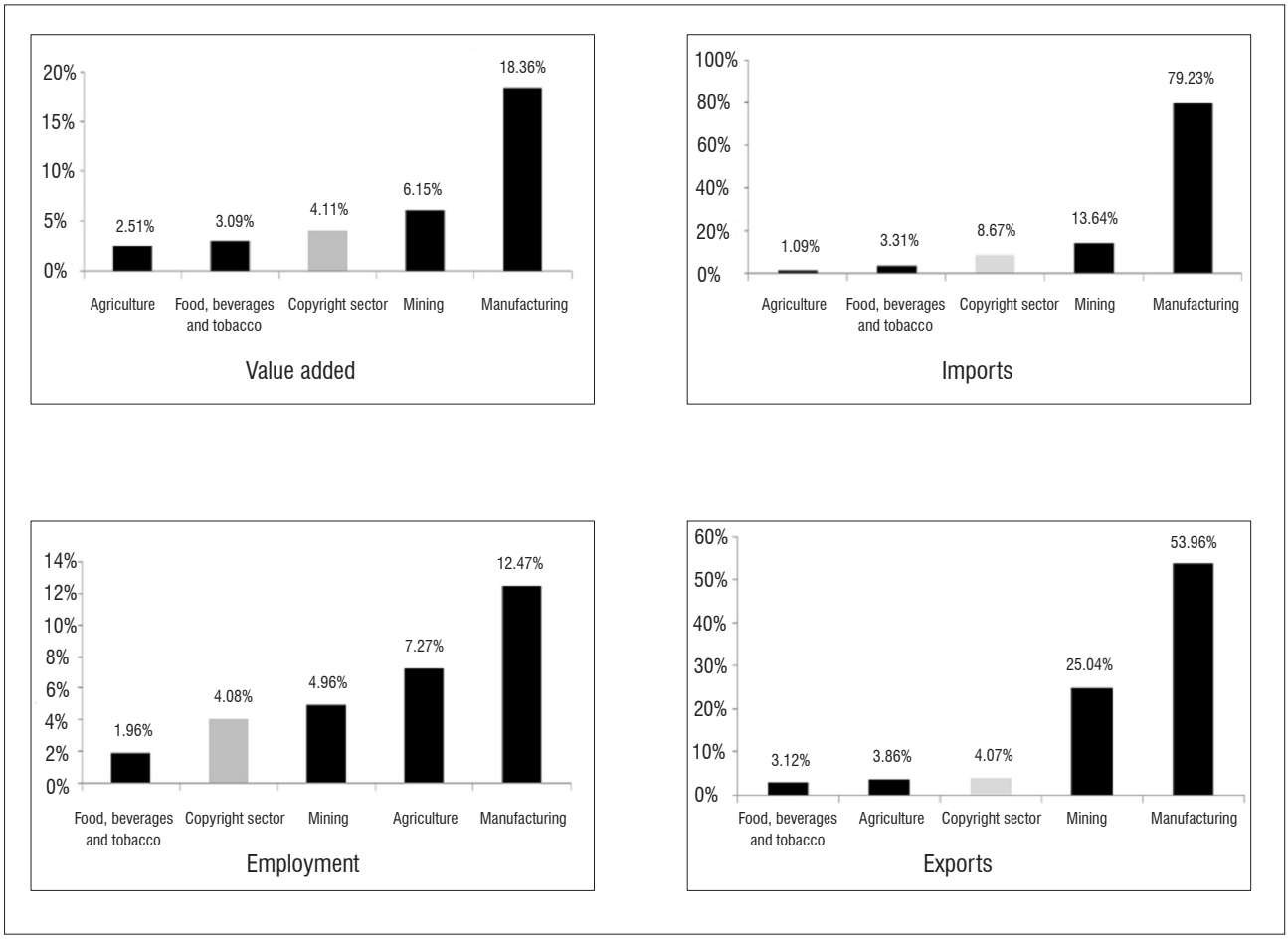
	Value added				Total
	Core	Interdependent	Partial	Non-dedicated	
1970–1979	28%	63%	36%	44%	37%
1980–1989	3%	20%	94%	18%	8%
1990–1999	17%	15%	-6%	31%	16%
2000–2008	24%	40%	24%	45%	33%
	Employment				**Total**
	Core	**Interdependent**	**Partial**	**Non-dedicated**	
1970–1979	33%	23%	29%	21%	33%
1980–1989	36%	16%	30%	10%	23%
1990–1999	51%	-5%	12%	4%	26%
2000–2008	30%	-6%	-13%	17%	21%
	Imports				**Total**
	Core	**Interdependent**	**Partial**	**Non-dedicated**	
1970–1979	9%	-28%	-31%	-20%	-7%
1980–1989	-17%	5%	-9%	49%	-5%
1990–1999	21%	231%	175%	27%	104%
2000–2008	-29%	72%	160%	8%	46%
	Exports				**Total**
	Core	**Interdependent**	**Partial**	**Non-dedicated**	
1970–1979	-55%	10%	93%	1%	15%
1980–1989	-14%	130%	101%	40%	84%
1990–1999	202%	97%	18%	100%	82%
2000–2008	108%	4%	23%	50%	31%

Table 3: Production-induced effect of copyright-based industries: Results of an input-output analysis of 2009

	Production-induced effect	Adjusted for copyright factors	Production-induced effect	Adjusted for copyright factors
Industry	**Outcome**		**Employment**	
Photography/software and databases/advertising	1.13%	0.11%	2.97%	0.30%
Communication	1.18%	0.07%	2.73%	0.16%
Crafts†	0.76%	0.32%	2.87%	1.21%
Film and television†	1.60%	1.60%	2.98%	2.98%
Footwear	2.08%	0.01%	6.23%	0.02%
Furniture	1.93%	0.19%	6.48%	0.65%
Glass and glass products	1.49%	0.01%	4.89%	0.03%
Computers and equipment/photocopiers	1.70%	0.06%	4.93%	0.17%
Other manufacturing	1.23%	0.12%	3.65%	0.37%
Paper and paper products	2.02%	0.50%	5.77%	1.44%
Printing, publishing and recorded media	1.79%	1.79%	5.17%	5.17%
Television, radio and communication equipment	1.64%	0.57%	4.75%	1.66%
Textiles	1.81%	0.01%	6.05%	0.02%
Transport and storage	1.19%	0.07%	3.02%	0.17%
Wearing apparel	1.59%	0.01%	6.19%	0.02%
Wholesale and retail trade	1.00%	0.06%	2.63%	0.15%
Total		**5.49%**		**14.52%**

Source: Adapted from WIPO[11]

†Denotes industries with extrapolated figures



Source: Adapted from WIPO[11]

Figure 5: Comparison of the contribution to the economy of copyright-based industries with those of the main economic sectors in South Africa in 2008.

Discussion

Our aim was to quantify the importance of copyright-based industries to the South African economy. The findings indicate that the contribution of these industries to value added, employment and trade should not be neglected: as expected they could not of course overpass the dominant sectors of the economy (mining and manufacturing) but they outperformed the 'agriculture' and 'food, beverages and tobacco' sectors. Our results are in agreement with those of Manfredi et al.[18] who showed that copyright-related industries play an important role in a country's economic growth and employment trends.

According to WIPO[11]:

> *The copyright-based industries are responsible for almost 4.11% of the total economy in terms of value added, with core copyright-based industries being the highest contributor (2.05%) and the non-dedicated copyright industries following with 1.29%. As far as employment is concerned, 4.08% of the workforce is employed in the copyright-based industries; the majority of which is employed in the core and non-dedicated copyright-based industries (2.31% and 1.03%). The interdependent copyright-based industries show a high contribution in the exports of the economy (2.77%) and an even higher contribution to the total imports (7.85%).*

Efforts have been coordinated and supported locally and internationally (primarily by WIPO) in quantifying the relevant effect on the economy. The economic structure and the promotion and incentives offered for the expansion and development of these industries are crucial factors. In addition, a shortage of relevant data hinders the country's already limited expertise in investigating the sectors based on copyright. Hence, it is suggested the relevant South African policy authorities and policymakers, such as the Department of Trade and Industry, monitor and publicise regularly the performance of the copyright-based industries and promote programmes for their development and growth.[19,20] Furthermore the country's funding agencies should support research related to copyright coefficients and the operations of the copyright industry.

Acknowledgements

This investigation was undertaken with the financial and intellectual support of WIPO. The article is based on a relevant WIPO report. We thank WIPO for their support and guidance.

Authors' contributions

The authors contributed equally to this study. R.I-L. provided expertise in quantitative analysis and A.P. provided expertise in policy implementation.

References

1. World Intellectual Property Organisation (WIPO). National studies on assessing the economic contribution of the copyright-based industries no. 1: The economic contribution of copyright-based industries in Singapore 2004. Singapore: WIPO; 2006.

2. Cvetkovski T. Global governance: Regulation of copyright law and policy in popular media copyright industries. In: Copyright and popular media. London: Palgrave; 2013. p. 53–88. https://doi.org/10.1057/9781137024602_2

3. Gani-Ikilama M. Copyright theory and a justificatory framework for creative autonomy in cultural industries. Queen Mary J Intelle. 2016;6(2):154–174. https://doi.org/10.4337/qmjip.2016.02.02

4. Chen L. China's creative industries: Copyright, social network markets and the business of culture in a digital age. New Media Soc. 2013;15(1):157–158. https://doi.org/10.1177/1461444812459453d

5. Scaria AG. Piracy in the Indian film industry: Copyright and cultural consonance. Cambridge: Cambridge University Press; 2014. https://doi.org/10.1017/CBO9781107588325

6. Yao JJ, Peng H. Copyright protection and cultural industry – Theoretical and empirical studies [article on the Internet]. c2014 [cited 2017 Mar 27]. Available from: http://citeweb.info/20142401529

7. Dobusch L, Schussler E. Copyright reform and business model innovation: Regulatory propaganda at German music industry conferences. Technol Forecast Soc. 2014;83:24–33. https://doi.org/10.1016/j.techfore.2013.01.009

8. Ericsson S. The recorded music industry and the emergence of online music distribution: Innovation in the absence of copyright (reform). George Wash Law Rev. 2011;79(6):1783–1813.

9. Nyariki D, Wasonga O, Otieno C, Ogadho E, Ikutwa C, Kithinji J. The economic contribution of copyright-based industries in Kenya. Nairobi: World Intellectual Property Organization; 2009.

10. Nicholson RD, Kawooya D. The impact of copyright on access to public information in African countries: A perspective from Uganda and South Africa. In: Proceedings of the World Library and Information Congress: 74th IFLA General Conference and Council; 2008 August 10–14; Quebec, Canada. Quebec: IFLA; 2008. p. 1–24.

11. World Intellectual Property Organization (WIPO). The economic contribution of copyright-based industries in South Africa. Pretoria: WIPO; 2011.

12. South African Department of Arts, Culture, Science and Technology (DACST). Cultural industries growth strategy. Pretoria: DACST; 1998.

13. World Intellectual Property Organization (WIPO). Guide on surveying the economic contribution of the copyright-based industries. Geneva: WIPO; 2003.

14. Siwek SE. The measurement of copyright industries: The US experience. Rev Econ Res Copyright Issues. 2004;1(1):17–25.

15. Media Group. The contribution of copyright and related rights to the European economy. Turku: Business Research and Development Centre, Turku School of Economics and Business Administration; 2003.

16. BSA-IDC. Sixth annual BSA-IDC global software: 08 Piracy study. Washington DC: Business Software Alliance; 2009. Available from: http://global.bsa.org/globalpiracy2008/studies/globalpiracy2008.pdf

17. Property Rights Alliance. The international property rights index [homepage on the Internet]. c2017 [cited 2017 Mar 27]. Available from: http://internationalpropertyrightsindex.org/

18. Manfredi S, Ricci F, Nappo F, Gilvari IS. Economic contribution of the intensive copyright industries: Theoretical and practical implications in Italy. Technol Innov Educ. 2016;2(1):1–17. https://doi.org/10.1186/s40660-015-0007-8

19. He T. What can we learn from Japanese anime industries? The differences between the domestic and oversea copyright protection strategies towards fan activities. Am J Comp Law. 2014;62:4. https://doi.org/10.5131/AJCL.2014.0029

20. Cheung M. Copyright challenges facing the website design industry: A survey with creative directors in Hong Kong. Des J. 2014;17(2):291–313. https://doi.org/10.2752/175630614X13915240576149

In the footsteps of Einstein, Sagan and Barnard: Identifying South Africa's most visible scientists

AUTHORS:
Marina Joubert[1]
Lars Guenther[1]

AFFILIATION:
[1]Centre for Research on Evaluation, Science and Technology (CREST), Stellenbosch University, Stellenbosch, South Africa

CORRESPONDENCE TO:
Marina Joubert

EMAIL:
marinajoubert@sun.ac.za

DATES:
Received: 03 Feb. 2017
Revised: 17 May 2017
Accepted: 06 July 2017

KEYWORDS:
public visibility; science communication; scientists and public engagement; science and its publics; science and society

HOW TO CITE:
Joubert M, Guenther L. In the footsteps of Einstein, Sagan and Barnard: Identifying South Africa's most visible scientists. S Afr J Sci. 2017;113(11/12), Art. #2017-0033, 9 pages. http://dx.doi.org/10.17159/sajs.2017/20170033

ARTICLE INCLUDES:
× Supplementary material
× Data set

FUNDING:
National Research Foundation (South Africa)

Highly visible scientists are increasingly recognised as influential leaders with a special role to play in making science part of mainstream society. Through consultation with a panel of 45 experts working at the science–media interface, we sought to identify the most visible scientists currently living and working in South Africa. In total, 211 scientists – less than 1% of the scientific workforce of the country – were identified as visible in the public sphere. The demographic profile and institutional spread of South Africa's visible scientists suggest that more should be done to increase the diversity of scientists who are publicly visible. Although only 8% of South Africans are white, 78% of the group of visible scientists were white, and 63% of the visible scientists were men. Only 17 black women were identified as publicly visible scientists. While visible scientists were identified at 42 different research institutions, more than half of the visible scientists were associated with just four universities. Recent controversies surrounding the two most visible South African scientists identified via this study, and the potential implications for fellow scientists' involvement in public engagement, are briefly discussed.

Significance:

- This is the first study to identify highly visible scientists in South Africa.
- The study has meaningful policy implications for mobilising scientists towards public science engagement.
- It is an important contribution towards the new public engagement framework of the Department of Science and Technology.

Introduction

A dominant trend in science over the past century is the move away from a closed system in which scientists worked largely in isolation, to a more transparent system in which scientists leave the proverbial ivory tower to engage with society.[1] Today, effective communication between scientists and diverse public audiences is recognised as an important characteristic of a modern, democratic knowledge society.[2] Consequently, the notion that scientists should engage with affected publics and debate the social implications of their work openly has become a central element of the moral economy of modern science. However, in a world where science has become increasingly competitive, contested and politicised[3], the borders between science and politics are blurring and science is more closely coupled with politics and mass media[4]. As a result, public communication of science has also become a tool to compete for public attention and political support[5] and scientists' motivation to seek public visibility may be complex and diverse.

The growing participation of scientists in public life has been accompanied by a growth in the presence of scientists in popular culture.[6] Iconic scientists who have shaped the societal discourse about science include people such as Charles Darwin, who was deemed a master of public relations[6] and Albert Einstein, *Time* magazine's person of the last century[7]. Towards the end of the 20th century, people like Stephen Jay Gould, Richard Feynman, Carl Sagan and Jane Goodall rose to fame, while present-day science popularisers such as Richard Dawkins, Stephen Hawking, Neil deGrasse Tyson, Brian Greene and David Attenborough continue to enjoy star status. As science continues to expand, it becomes harder for scientists to be visible in public life and the relative invisibility of scientists has emerged as a major concern.

Scientists' academic visibility within their fields of research depends on scholarly publication and citation rates, but public visibility depends on media exposure[8], which may result from a combination of academic work and involvement in debates and activities outside science. Scientists' active participation in public science communication (for example, presenting popular talks at schools and science centres) is not sufficient to ensure a high public profile. Public visibility requires an amplification of the individual scientist's views and voice – something that would be virtually impossible to achieve without the strategic use of traditional and digital media platforms.[6,9] Because of the decisive role of the media in achieving public visibility, we identified the most visible scientists in South Africa with the help of a panel of science–media experts, that is, science writers and researchers in the field of science communication and/or media studies.

The case of Carl Sagan is frequently cited as an example of how a high public profile can detract from a scientific career. It is widely believed that Sagan's nomination to the US National Academy of Sciences was voted down by his peers, because they did not approve of his celebrity status.[6] Subsequently, scientists' fears that public visibility could penalise them career-wise became known as the 'Sagan effect'. Interestingly, Sagan himself did not perceive that he was penalised by his peers as a result of his high public profile.[8]

Despite lingering ambivalence amongst some scientists about the desirability of achieving a high public profile[10], media interactions – leading to public visibility – have become an integral part of the role of being a scientist and are particularly expected of scientists in leadership roles[11]. Because of their ability to spread their ideas, influence decision-makers and promote scientific culture, high-profile scientists are considered by some as the new academic elite[6] and increasingly recognised as powerful socio-political influencers and opinion leaders who

have power within and outside the science arena.[12] Furthermore, high-profile scientists have been shown to outperform their less visible peers in terms of scientific productivity[13,14] and their media visibility may well boost their scientific impact[15,16]. That is why we had the goal of identifying and locating the most visible scientists in South Africa, with a special focus on their population group, gender, age, employment sector and research fields. This approach is beneficial in terms of identifying what kind of scientists become publicly visible in South Africa and testing whether the publicly visible scientists, as identified in the current study, are representative of the broad South African society. Furthermore, the study has the potential to generate recommendations for increasing the visibility of scientists in South African society.

While the scientific community may have become more accepting of public visibility, some scientists remain concerned about the potentially damaging effects of a high public profile on their scientific reputations.[10,17] Their concerns are exacerbated by apprehensions about the risks presented by social media.[18] Furthermore, the rise of public relations approaches in institutional science communication[19] has resulted in some criticism of the efforts of some institutions and scientists to gain attention for their work[5]. The inevitable tensions between increasing expectations to engage with public audiences and lingering reservations about the potential consequences of doing so, sustain scientists' ambivalence about public visibility.[10] Given these conflicting prospects in terms of public visibility, it becomes even more important to understand how scientists themselves experience and respond to opportunities and demands to communicate publicly about their research.

The South African context for public science communication

During the Portuguese colonial rule in Brazil from the 16th to 18th centuries, science was typically suppressed and only the rulers had access to scientific knowledge as a consequence of their links with Europe.[20] This was also the case during British colonial rule in South Africa (1795–1910). The country's science system continued to serve mostly its white population and government interests[21] during the racially segregated apartheid regime (1948–1994), thereby continuing to isolate the majority of the country's citizens from science and suppressing the communication of scientific ideas[22]. Despite international isolation, boycotts, sanctions and travel restrictions during the apartheid years, the South African science base continued to strengthen and the country developed advanced facilities and expertise in fields such as geology, mining, energy, nuclear science, space science and military science, as well as in agriculture and veterinary sciences.[23]

During the apartheid years, only one South African scientist became a household name locally and abroad: Christiaan Barnard. Christiaan Barnard (1922–2001) shot to global fame after he performed the first human-to-human heart transplant in Cape Town on 3 December 1967. Barnard's article describing the surgery was published within 4 weeks of the event[24] and became one of the most cited articles in the field of cardiovascular medicine[25]. The dramatic nature of this medical milestone and the subsequent events captured the attention of the world and they were front page news for some time. It is likely that Barnard's youth and charisma enticed ongoing media attention, but it has also been suggested that politicians exploited him to improve South Africa's image at a time when the country was politically and socially isolated.[25]

South Africa has delivered other celebrated academics who were highly regarded within science, but they – arguably – never became household names. Names that come to mind include the visionary veterinary researcher Sir Arnold Theiler and the eminent and much-loved palaeoanthropologist Phillip V. Tobias.

The arrival of democracy in 1994 changed South African society fundamentally. In this new dispensation, science and technology were seen as essential instruments for economic growth and social development.[26] In addition to a new political and social landscape, there was an expectation that science should also be democratised. The science base at the time of transition to democracy was based on only about 10% of the population. The country now had a much larger pool of talent on which to draw, but developing this expertise would require extensive educational reform, as well as intensive teaching and mentoring in order to correct the injustices and imbalances of the past.[27] Consequently, new government policy emphasises the need for more effective engagement between science and society, and the government is urging the country's scientists to become actively and visibly involved in public science communication in order to make their work accessible, meaningful and relevant to all South Africans, in particular historically disadvantaged communities.[28] These calls are in line with increasing demands on scientists to engage pro-actively with a wide range of public and policy audiences, including entering into meaningful dialogue with society about uncertainty in science and the ethical implications of cutting-edge research.[29]

The Year of Science and Technology 1998 – or 'Yeast 98' as it was popularly known – was the first nationwide science communication initiative in democratic South Africa. This public communication campaign consisted of exhibitions, popular science talks and interactive workshops that moved around the country. Subsequently, the government has organised annual science and technology weeks and contributed funding to various science centres and science festivals. In 2002, the Department of Science and Technology established the South African Agency for Science and Technology Advancement (SAASTA) as a business unit of the National Research Foundation, with the specific aim to advance public awareness, appreciation and engagement of science, engineering and technology in South Africa. Early in 2015, the Department of Science and Technology announced a new strategic framework for public science engagement[28], which is intended to coordinate and encourage science promotion, communication and engagement activities at a national level. The stated objectives of the framework blend promotional aims (popularising science to awaken interest and global profiling of South African science) and engagement goals (developing a critical public that participates actively in the national discourse on science and technology).

Achieving the government's ambitious public engagement goals will require not only the support of South Africa's institutional science system, but also the participation of as many individual scientists as possible. However, in order to design effective strategies and support structures to mobilise scientists in terms of public science communication, we must understand the motivations and barriers that influence their willingness and ability to get involved.

Several scholars have emphasised the significance of understanding how scientists themselves perceive public communication and their roles in public life as a first step in developing effective public science engagement strategies.[30,31] While the public communication of science, including the factors that influence scientists in terms of their communication with lay audiences, has attracted considerable research interest over the last few decades, local studies were mostly limited to explorations of public science literacy and attitudes to science.[32] Only two studies investigated interactions between South African scientists and journalists[33,34] and both highlighted gaps in understanding and skills deficits on the sides of both journalists and scientists.

Currently, we still know very little about the motivations and barriers that influence South African scientists' public communication behaviour, and how they perceive the benefits and risks associated with a high public profile. It is hoped that future research will help to inform new policies and support structures aimed at encouraging and incentivising scientists towards more visible participation in public life. The current study is a first step towards achieving this goal. Consequently, the research question steering the current study is: Who and where are the visible scientists in South Africa?

Background information about these scientists – including their demographic profiles, fields of research and institutional affiliations – provides new perspectives on the factors that determine scientists' visibility in the public sphere within South Africa's unique science arena and socio-political context.

Methodology

For the purpose of this study, public science communication is defined as the communication of scientific information – by scientists – to people not involved with research in the same field. Visible scientists were identified as those scientists who were considered to be known to the South African public – at least to some extent – by one or more members of a panel of individuals working at the interface between science and the media.

A number of factors – including age[35], gender[36], position[37], discipline[38] and organisational culture[39] – have been shown to influence if and how scientists communicate with public audiences. As such, the aim of the current study was to identify publicly visible scientists in South Africa, along with relevant background information about them that might help explain their public visibility.

After considering various options for identifying publicly visible scientists (such as selecting recipients of science engagement awards, choosing scientists based on the frequency of media appearances or doing a public opinion survey), asking a panel of science–media experts (similar to the approach followed by Rae Goodell in her landmark study of visible scientists in 1975[8]) was deemed to be the most feasible and effective. Goodell points out that using a panel of respondents – each with their own experiences of scientists interacting with the public – allows the researcher to draw indirectly on a number of indicators of visibility.

Based on the first authors' experience of 25 years in the South African science communication arena, panel members were selected to represent local expertise in science journalism, popular science writing, public science engagement, and research in the field of science communication. To avoid bias in favour of visible scientists at a specific research organisation, no corporate communicators associated with a specific institution (university or science council) were included in the panel; this exclusion was not applied to researchers based at institutions.

Potential panel members were contacted via email (with reminders via email and social media) in order to ask them to help identify scientists who they perceived to have achieved some level of public visibility in South Africa. In line with the approach used by Goodell[8], the request was kept brief and simple; respondents were asked to be liberal in their definition of a scientist. Following a short introduction to the research project, one simple question was asked: 'Please write down the names of about five to ten scientists – currently living and working in South Africa – who you consider to be publicly visible.' Respondents were asked to name a maximum of 10 scientists for two reasons: (1) to ensure that it would not be a time-consuming task, and also (2) to encourage them to submit names of only the scientists who were well known in the public sphere.

Of the 63 experts who were approached, 45 (71%) responded to the request by sending names of scientists who they perceived to be publicly visible in South Africa. Of these 45 respondents, 41 are white (4 are black) and 29 are women (16 are men). The industry sectors represented by panel members are reflected in Table 1.

Correspondence took place during November and December 2016. From the responses, names of South African born scientists who were no longer in the country at the time of the study (such as Elon Musk and Mark Shuttleworth) were deleted, as were the names of deceased scientists (such as Phillip Tobias). From the names provided by the respondents, further desktop research was undertaken to determine the population group, gender and research fields of the group of 211 scientists identified as publicly visible, as well as to determine the age profile of the 'top 18' scientists (i.e. the scientists mentioned most often by the respondents).

Ethical clearance for this study was obtained from the Research Ethics Committee: Human Research (Humanities) at Stellenbosch University (reference number SU-HSD-004069).

Results

The number of visible scientists in South Africa

According to 2014/2015 statistics, South Africa has 25 300 researchers (*excluding* doctoral students and postdoctoral fellows at higher education institutions).[40] Researchers are defined as professionals engaged in the conception or creation of new knowledge, products, processes, methods and systems, and in the management of the projects concerned.[40] (The terms 'researcher' and 'scientist' are used interchangeably in this article.) Of these 25 300 researchers who make up the scientific workforce, only 211 scientists (less than 1%) were identified as being publicly visible by the 45 respondents who constituted the media panel for this study. In total, 367 names were suggested, representing 211 individuals once duplicate mentions were removed. Given that the maximum number of names expected if each respondent listed 10 unique names would be 450, the list of 367 names represented 82% of the expected maximum.

Only 46 of the 211 scientists were mentioned by two or more panel members. This means that 165 scientists (78% of the total group) were mentioned by only one person on the media panel, suggesting that most of the visible scientists identified in the current study are not genuinely publicly visible, but rather known as a result of limited media exposure, with possibly only one journalist. Professor Lee Berger – a palaeoanthropologist at the University of the Witwatersrand – emerged as the most visible scientist in South Africa with 27 mentions.

The 'top 18' – i.e. scientists who were mentioned by four or more panel members – are listed in Table 2, along with their research fields and institutional affiliations. This group of 18 scientists can therefore be defined as having achieved a significant level of public visibility.

South Africa's visible scientists per population group

While white scientists constitute 59% of the current scientific workforce of the country[40], 164 of the visible scientists in this study (or 78% of the visible group) were white. Black researchers (consisting of African, coloured and Indian researchers) constitute 42% of the scientific workforce.[40] Only 47 black researchers (22% of the group of 211 visible researchers) were identified as being publicly visible.

The 164 white scientists identified as publicly visible represent 1% of the white scientists in South Africa, while the 47 black scientists represent only 0.4% of the black scientists in the country. The 18 most visible scientists (Table 2) are made up of 11 white and 7 black scientists.

The gender balance of South Africa's visible scientists

The male:female ratio in South Africa's scientific workforce[40] is 56% male to 44% female. Men outnumber women in the scientific workforce, but were even more dominant amongst the group of 211 visible scientists identified in the current study, with 133 (63%) men and 78 (37%) women. Similarly, 11 (61%) of the 18 most visible scientists (as listed in Table 2) were men. There were only 17 black women (8% of the visible group) amongst the visible scientists.

Age and seniority profile of the most visible scientists

All of the most visible scientists identified in this study (i.e. the top 18 scientists mentioned by four or more of the respondents) were born between 1947 and 1971, meaning that – on their birthday in 2016 – their ages ranged between 45 and 69 years, with an average age of 52 years. Only two people in this group of scientists were younger than 50 years at the time.

In terms of seniority, 14 of the 18 most visible scientists (77%) were full professors, while 1 (who was also the youngest in the group) was an associate professor. Two of the remaining three scientists held a PhD, but were not employed in the higher education sector. Only one of the top 18 most visible scientists did not have a PhD.

Table 1: Science–media panel respondents ($n=45$)

Name and surname	Affiliation	Industry sector
Duncan Alfreds	Media24	Online (general news website)
Sheree Bega	*Saturday Star*	Print media (newspaper)
Sue Blaine	*Financial Mail*	Print media (magazine)
Nana Boaduo	National Research Foundation	Science communication (agency)
Val Boje	*Pretoria News*	Print media (newspaper)
Elsabé Brits	Netwerk24	Print and online news
Tony Carnie	*The Mercury*	Print media (newspaper)
Katharine Child	*The Times*	Print media (newspaper)
Lesley Cowling	University of the Witwatersrand	Research
Marize de Klerk	Video News 247	Online (video news)
Izak du Plessis	SABC	Broadcast (radio)
Harry Dugmore	Rhodes University	Research
Engela Duvenage	Freelance	Science communicator
Alex Eliseev	EWN	Broadcast (radio)
Amelia Genis	*Landbouweekblad*	Print media (magazine)
Yolandi Groenewald	City Press	Print media (newspaper)
Daryl Ilbury	Freelance	Science communicator
Natasha Joseph	*The Conversation Africa*	Online (research news and opinion)
Tamar Kahn	*Business Day*	Print media (newspaper)
Lia Labuschagne	Freelance	Science communicator
Ruda Landman	Freelance	Freelance (television)
Steven Lang	Freelance	Freelance (print media)
Tony Lelliot	University of the Witwatersrand	Research
Janice Limson	Rhodes University	Research
Stephan Lombard	Cape Talk Radio	Broadcast (radio)
Munyaradzi Makoni	Freelance	Online and print media
Mia Malan	*Mail & Guardian*	Print media (newspaper)
Izak Minnaar	SABC	Online (news)
Anina Mumm	ScienceLink	Science communicator
Thabiso Nkone	National Research Foundation	Research management
Linda Nordling	Freelance	Print media
Joanne Riley	SAASTA	Research management
Freek Robinson	*kykNET*	Broadcast (television)
Ina Roos	Freelance	Print media
Marika Sboros	Freelance	Print media
Elna Schütz	The Wits Radio Academy	Broadcast (radio)
Mandi Smallhorne	Freelance	Print and online media
Lynne Smit	Hippo Communications	Science communicator
Elise Tempelhoff	*Beeld*	Print media (newspaper)
Anso Thom	Health-e/Section27	Online (health news)
Irma Venter	*Engineering News*	Print media (magazine)
Derek Watts	*Carte Blanche*	Broadcast (television)
Mandy Wiener	Freelance	Broadcast (radio)
Sarah Wild	Freelance	Print and online media
John Yeld	Freelance	Print media

Respondents were advised that their names would be published unless they objected; there were no objections.

Table 2: Publicly visible scientists mentioned by four or more members of the science–media panel

Scientist	Research field	Institution	Mentions
Professor Lee Berger	Natural sciences (palaeoanthropology)	University of the Witwatersrand and National Geographic explorer-in-residence	27
Professor Tim Noakes	Health sciences (exercise, sports science, nutrition)	University of Cape Town (Emeritus Professor) and The Noakes Foundation	14
Professor Glenda Gray	Health sciences (perinatal HIV care)	Medical Research Council	12
Professor Bob Scholes	Natural sciences (ecology)	University of the Witwatersrand	10
Professor Salim Abdool-Karim	Health sciences (epidemiology and infectious diseases)	University of KwaZulu-Natal and Centre for the AIDS Programme of Research in South Africa (CAPRISA)	10
Professor Anusuya Chinsamy-Turan	Natural sciences (palaeobiology)	University of Cape Town	8
Professor Himla Soodyall	Natural sciences (population genetics and molecular anthropology)	University of the Witwatersrand and National Health Laboratory Service	8
Doctor Bernie Fanaroff	Physical sciences (astronomy/astrophysics)	South African Square Kilometre Array Project (National Research Foundation)	7
Professor Anthony Turton	Natural sciences (water resource management)	University of the Free State and TouchStone Resources	6
Professor Linda-Gail Bekker	Health sciences (HIV, infectious disease and molecular medicine)	University of Cape Town (Desmond Tutu HIV Centre)	5
Professor Tebello Nyokong	Physical sciences (medicinal chemistry and nanotechnology)	Rhodes University	5
Dave Pepler	Natural sciences (ecology and conservation)	Academy for Environmental Leadership	4
Professor Bongani Mayosi	Health sciences (dean of faculty)	University of Cape Town	4
Professor Bruce Rubidge	Natural sciences (palaeontology)	University of the Witwatersrand	4
Professor Kelly Chibale	Physical sciences (drug discovery and medicinal chemistry)	University of Cape Town	4
Professor Francis Thackeray	Natural sciences (palaeontology)	University of the Witwatersrand	4
Professor Jill Farrant	Natural sciences (drought tolerance in plants)	University of Cape Town	4
Professor Nox Makunga	Natural sciences (medicinal plants)	Stellenbosch University	4

South Africa's visible scientists per employment sector

A breakdown of the institutions at which the visible scientists work (Table 3) revealed that just more than half (109, or 52%) of the 211 publicly visible scientists identified in this study were working at just four universities. A total of 157 out of 211 (74%) were employed at just 10 institutions. The University of Cape Town was home to most of the high-profile scientists in South Africa, followed by the University of the Witwatersrand, University of Pretoria and Stellenbosch University. The rest of the visible scientists were spread over 42 other research institutions, of which 27 institutions were represented by only one scientist. Notably, there are many more research organisations (science councils and universities) in the country from which not a single scientist was identified as being publicly visible.

In total, 160 of the 211 visible scientists (76%) in the country were employed in the higher education sector, with 22 (10%) employed in science councils and 12 (6%) in the not-for-profit sector. There were relatively few visible scientists in industry (only 7, or 3%), and even fewer (only 5, or 2%) in government. Five of the visible scientists were retired and no longer affiliated to a specific institution.

Table 3: The top ten South African institutions in terms of the number of publicly visible scientists

Institution	Number of visible scientists
University of Cape Town	37
University of the Witwatersrand	34
University of Pretoria	20
Stellenbosch University	17
Rhodes University	9
University of KwaZulu-Natal	9
University of Johannesburg	9
National Research Foundation (Square Kilometre Array South Africa and the South African Astronomical Observatory)	7
North-West University	7
University of the Free State	7

Table 4: Disciplinary spread of publicly visible scientists in South Africa

Broad research field	Number of visible scientists
Natural, biological and agricultural sciences (including animal and plant studies, microbiology, conservation, marine biology, biotechnology, genetics, agriculture, food security, palaeontology and palaeoanthropology)	64
Social sciences, arts and humanities (including education, communication, media studies, history, economics, political science, law)	50
Health sciences (including HIV/Aids, disease, public health, nutrition, sports science and private medical practice)	38
Physical sciences (including mathematics, chemistry, physics, nanotechnology, astronomy, astrophysics and space science)	31
Environmental sciences (including climate sciences, earth sciences, waste, water and pollution)	16
Engineering (including energy, materials science, chemical engineering, infrastructure, electronics)	12

South Africa's visible scientists per broad research field

Table 4 provides a breakdown of the visible scientists in South Africa according to broad research field. Most of the visible researchers in South Africa, as identified in this study, worked in the natural sciences and it is interesting to note that this group included eight researchers in the field of palaeontology. The combined disciplines of social sciences, humanities and arts represented the second biggest group of visible scientists, followed by health sciences, physical sciences and environmental sciences. Only 12 engineers (6%) were identified as publicly visible.

Discussion

Too few black and women scientists in the public eye

Only 211 of the 25 300 South African researchers[40] – less than 1% – were identified as being publically visible within South Africa. Within this group, black and female scientists were proportionally underrepresented.

Black scientists make up 42% of the researchers in South Africa, but only 33% of the visible scientists in this study. Similarly, women constitute 44% of the research workforce, but only 34% of the visible group.

When these figures are seen in the context of the overall population demographics of the country, the underrepresentation of black and female scientists is even more prominent. According to recent population estimates from Statistics South Africa[41], the total South African population stands at 55.9 million people, of whom 92% are black and 51% are female. White people constitute only 8% of the overall South African population, but 78% of the visible scientists in this study (164 out of 211) were white.

Given the political past of South Africa, one could argue that it is encouraging that there are 47 black scientists (including 17 women) deemed to be publicly visible at the present moment in South Africa. A deeper understanding of their experiences of public communication of science could reveal the best way forward to increase the participation of black scientists in public communication about their research.

In terms of the effect of gender on careers in science, the dominance of men at higher levels in the academic hierarchy is well documented, as is the evidence for a general structural bias against women in science[42], while female scientists in Africa face a particularly dire situation[43]. Women also face particular barriers in terms of getting involved in public science engagement and may even be advised to avoid these activities or risk not being taken seriously by their male peers.[44] There is also evidence that active involvement in public engagement is generally valued in a man's portfolio, but criticised when part of a woman's portfolio.[45] Given the persistent stigmatising of women who are actively involved in public science communication, it is no surprise that some female scientists avoid or downplay personal involvement in these activities and understandable that nearly two thirds (133 out of 211) of the visible scientists identified in this study were men. In contrast, some studies show that, despite the normative sanction from their peers, female scientists are significantly more involved in reaching out to external audiences compared with their male colleagues.[31]

The dominance of male scientists in public life is not unique to South Africa. For example, there were only two women – Margaret Mead and Jane Goodall – in the list of the 20 most visible scientists identified by Goodell[8]. Similarly, Fahy[6] included only one woman – Susan Greenfield – in his chronicle of the lives of high-profile scientists of today.

However, in South Africa, increasing the gender and racial diversity of the science workforce (i.e. attracting young women and black youth to research careers), is an important objective of public science communication. Publicly visible black and female scientists can act as role models for young people and help to shatter the racial and gender stereotypes in science. It is therefore important to understand how local black and female scientists respond to demands for increased public science engagement, and what specific factors encourage or constrain their involvement.

Public visibility increases with seniority

Research has shown that journalists often prefer to interview scientists who are more senior and likely to be in influential positions and regarded as leaders in their fields, and therefore it takes time for scientists to achieve public visibility; also senior scientists are more likely to engage with the public and be better able to deal with the potentially negative responses that may result from public visibility from time to time.[8,35,38,46] This contention is also evident in the age profile of the 18 most visible scientists identified in this study who were between 45 and 69 years old, with an average age of 52 (and only two researchers younger than 50). These findings are in line with the higher levels of public communication involvement of more senior researchers who have been demonstrated in, for example, Switzerland[36] and Argentina[37]. It has also been suggested that scientists should earn a scientific reputation before venturing out into the public arena.[8] It is therefore not surprising that 14 of the 18 most visible scientists were full professors, which allows them the credibility and protection of a high standing in the academic world.

The effect of organisational culture on public visibility

In today's global research arena in which universities increasingly have to justify research spending and compete for the attention of funders and partners, institutions are looking for ways to demonstrate their social relevance and impact via the public engagement activities of their researchers. Consequently, organisational culture, policies, norms, reward structures, institutional expectations and the quality and availability of communication support services have been shown to influence public visibility of researchers.[47,48] Scientists also strive for recognition within their institutions and may pursue public and media visibility in order to establish and boost their own reputations.[19] While organisational culture can support scientists' efforts to engage external audiences, institutional constraints mean that scientists may also be penalised for their efforts to become publicly visible.[1,39]

In this context, it is interesting to note that more than half of the 211 publicly visible scientists identified via this study were employed at just four South African universities. These universities – Cape Town, Witwatersrand, Pretoria and Stellenbosch – are all research-intensive universities ranked amongst the best higher education institutions in the country.[49] It is reasonable to assume that the organisational culture and policies, as well as the public relations support, that are available at these top universities, play a role in helping the scientists who work there to achieve and sustain higher public profiles. Apart from these four universities, the rest of the visible scientists in the country were thinly spread across the science sector, with many universities and other research organisations featuring just one publicly visible scientist. There were also several universities and science councils, as well as national and provincial government departments, missing from the list of institutions hosting visible scientists, meaning that none of the 45 media panel members mentioned a visible scientist at these institutions.

The effect of discipline on public visibility

There are many reasons why scientists' efforts to communicate with public audiences about their research may be affected by their own field of research. For example, the public tends to be more interested in topics that are close to everyday life and resonate with human interest.[8] Also, the esoteric nature of natural and physical sciences, and the highly codified language used in these fields, make it challenging to present new ideas to lay audiences.[13,50] Researchers working in some fields may perceive a strong moral duty to make their work publicly accessible, while other disciplines are governed by more restrictive norms.[31]

The number of visible South African scientists per broad scientific field, as identified in the current study, were in line with these findings and also reflected the scientific opportunities and priorities of the country. For example, given the country's rich biodiversity[51] and unique fossil heritage[52], it is not surprising that there were many biologists and palaeontologists amongst the visible scientists (Table 4). Similarly, given the local health challenges[53], it is understandable that HIV, tuberculosis and public health dominate as areas of expertise amongst the publicly visible health researchers. The importance and relevance of social scientists in a developing country context is underlined by the fact that nearly a quarter of the publicly visible scientists were social science scholars working on topics such as politics, economics, gender studies, communication, criminology, violence, trauma and reconciliation.

Amongst the 18 most visible scientists in the country, more than half (10) were from the natural sciences (of which 4 were in fields related to palaeontology), 5 were in health sciences, and 3 in physical sciences. Notably, there were no social scientists or engineers in this group of the most visible scientists. This finding indicates that scientists in these fields have therefore not achieved a similar level of visibility compared with the best-known experts in the country in fields such as palaeontology, climate change and HIV/Aids research.

Berger and Noakes – blending science celebrity and notoriety

Goodell[8] reflects on the discomfort that some scientists experience when peers or colleagues attain high public profiles, including that they are sometimes seen as irritating and even hazardous because of their tendency to break old rules of protocol in the scientific profession, question old ethics and defy old standards of conduct. She adds that scientists are often concerned that high-profile colleagues will mislead the public when they speak outside their areas of expertise – as they often do. These observations about visible scientists – and the ways in which other scientists respond – are reminiscent of recent controversies that have surrounded the two most visible scientists identified in this study – Lee Berger and Tim Noakes.

There is little doubt that controversy is a catalyst for public visibility[8] and that visibility in the mass media feeds further visibility via the feedback loops of media attention[54] and reciprocal intensification[39]. Lee Berger and Tim Noakes have both experienced considerable public and peer criticism that has certainly boosted their public visibility. Both of them are acclaimed, senior scientists: Lee Berger is rated as an 'internationally acclaimed researcher', and Tim Noakes as a 'leading international researcher' according to the evaluation and rating system of the National Research Foundation (as at 14 December 2016). Seniority and status have been shown to help scientists cope with critical storms that may result from high media visibility.[8,46]

When, towards the end of 2015, Lee Berger went public with his theory that *Homo naledi* was a human ancestor that in all likelihood deliberately buried its dead[55], the discovery became mired in controversy and elicited societal and scientific criticism[56]. Amongst other things, Berger was accused of rushing his research to please the media[57] and even of promoting racist pseudoscience intent on showing that Africans were sub-human[58]. In a press conference on 9 May 2017, Berger and his research team not only announced that many more of these fossils have been discovered, but also dated them and presented further claims that the species could have shared cognitive traits with modern humans, as their hands were capable of making and manipulating tools.

Tim Noakes' advocacy in favour of a diet low in carbohydrates and high in fat[59] has resulted in an outright diet war between passionate supporters, mostly members of the public, and fierce critics, who were mostly in the scientific community. Fellow scientists have challenged the scientific basis of the low-carbohydrate-high-fat diet[60], while colleagues at the University of Cape Town have accused him of making outrageous and unproven claims and have distanced themselves from his dietary recommendations[61].

Two South African science journalists – Sarah Wild and Alex Eliseev – weighed in on the Berger and Noakes controversies, and how they have played out in the public sphere. Wild[62] claims that so-called 'rock star scientists' – such as Berger and Noakes – are populists and that their style of communication – oversimplifications, appeals to emotions and anecdotes – threatens public trust in science (should they later be proven wrong). Eliseev[63] disagrees; he posits that to get people interested in science, you have no choice but to simplify. Furthermore, scientists routinely disagree and will debate fossil finds and diets for years to come. Eliseev describes Berger as one of those rare scientists with flare and personality – the X-factor – with a talent to make science exciting and entertaining. He sees no problem with Berger's ability to use the media to his advantage, adding that it helps to fund research and has other positive spin-offs for society.

Like Berger and Noakes, many high-profile scientists have on occasion experienced damaging criticism from peers. While their strong track record may help them to weather these storms, they are not oblivious to these attacks on their credibility, and often fear the criticism from their peers.[8] This effect is evident in responses from Noakes after he was found not guilty of unprofessional conduct, following a drawn out disciplinary hearing by the Health Professions Council of South Africa. Commenting on the letter by University of Cape Town academics, Noakes told a journalist[64]: 'What that letter did was to isolate me from my university and my faculty; from the university unit that I had started and from the Sports Science Institute of South Africa which I had also helped found'. About the drawn-out hearing itself, he said: 'My wife and I simply could not escape it – we spoke about it endlessly, wondering why it had come about; how we might have avoided it; how heartless were the people who were driving it and we even wondered at times whether our lives might be in danger.'

Controversies like these – characterised by an intense focus on a specific individual – may discourage other South African scientists from going public with their research findings. It is therefore important to understand how visible scientists cope with and respond to controversy, and to discover what lessons other scientists can learn from their experiences.

Study limitations and suggested future research

The present study has important limitations. To identify visible scientists in South Africa, only experts working at the science–media interface were considered, only 45 of them responded to the request, and the respondent group was predominantly white. In addition, these experts were asked to name only between 5 and 10 scientists. There might be other respondent groups to consider that could extend the number of

visible scientists in South Africa. Furthermore, only names of visible scientists were gathered; the background data were sourced from desktop research. These findings lead to interesting new research questions. For instance, it would also be prudent to investigate the communication culture and support structures at the four universities that are home to more than half of the publicly visible scientists identified in this study, in order to elucidate the successful communication drivers that could possibly be adopted by other research organisations across the country. A better understanding of the communication behaviour and experiences of high-profile scientists who have endured intense public and peer scrutiny – as is the case with the two most visible scientists in this study – will further aid our understanding of the factors that influence and shape the interaction between scientists and their diverse publics in a country like South Africa.

Conclusion

Despite the above-mentioned limitations, the present study has important implications. Given the science transformation goals of the country – especially in terms of attracting more black youth into science and demonstrating the social relevance of research – it is of importance to increase the visibility of black scientists and to raise the profile of black intellectuals in the country. To achieve these goals, a new generation of black scientists must be equipped with the confidence and skills to become publicly visible via mass media platforms. Similarly, more female scientists must be mobilised, motivated and incentivised to communicate and engage with the broad South African society. Therefore, in order to reach the goals outlined in its ambitious public engagement framework, the South African government will need to broaden the base of scientists who become visible in the public eye. Investigations into the profiles and experiences of high-profile scientists highlight the factors that influence public visibility and help to inform future policies designed to support public science communication.

Acknowledgements

We thank the National Research Foundation of South Africa for funding (grant number 93097).

Authors' contributions

M.J. was responsible for the research. M.J. and L.G. collaborated on the processing, presentation and discussion of the findings.

References

1. Bucchi M, Trench B. Science communication research: themes and challenges. In: Bucchi M, Trench B, editors. Routledge handbook of public communication of science and technology. New York: Routledge; 2014. p. 1–14.
2. Horst MA. Field of expertise, the organization, or science itself? Scientists' perception of representing research in public communication. Sci Commun. 2013;35(6):758–779. http://dx.doi.org/10.1177/1075547013487513
3. Scheufele DA. Science communication as political communication. Proc Natl Acad Sci USA. 2014;111(Suppl. 4):13585–13592. http://dx.doi.org/doi:10.1073/pnas.1317516111
4. Weingart P. The moment of truth for science. EMBO Rep. 2002;3(8):703–706. http://dx.doi.org/10.1093/embo-reports/kvf165
5. Weingart P, Guenther L. Science communication and the issue of trust. JCOM. 2016;15(5), C01, 11 pages.
6. Fahy D. The new celebrity scientists. New York: Rowman & Littlefield; 2014.
7. Golden F. Albert Einstein: Person of the century. Time Magazine. 1999;154(27):34–37.
8. Goodell RS. The visible scientists [dissertation]. New York: Stanford University; 1975.
9. Fahy D, Lewenstein B. Scientists in popular culture: The making of celebrities. In: Bucchi M, Trench B, editors. Routledge handbook of public communication of science and technology. New York: Routledge; 2014. p. 83–96.
10. Rödder S. The ambivalence of visible scientists. In: Rödder S, Franzen M, Weingart P, editors. The sciences' media connection–public communication and its repercussions. Dordrecht: Springer; 2012. p. 155–178.
11. Peters HP. Gap between science and media revisited: Scientists as public communicators. Proc Natl Acad Sci USA. 2013;110(Suppl. 3):14102–14109. http://dx.doi.org/doi:10.1073/pnas.1212745110
12. Baron N. Stand up for science. Nature. 2010;468:1032–1033. http://dx.doi.org/10.1038/4681032a
13. Bentley P, Kyvik S. Academic staff and public communication: A survey of popular science publishing across 13 countries. Public Underst Sci. 2011;20(1):48–63. https://doi.org/10.1177/0963662510384461
14. Jensen P. A statistical picture of popularization activities and their evolutions in France. Public Underst Sci. 2011;20(1):26–36. http://dx.doi.org/10.1177/0963662510383632
15. Liang X, Su LY, Yeo SK, Scheufele DA, Brossard D, Xenos M, et al. Building buzz: (Scientists) communicating science in new media environments. Journalism Mass Comm. 2014;91(4):772–791. http://dx.doi.org/10.1177/1077699014550092
16. Phillips DP, Kanter EJ, Bednarczyk B, Tastad PL. Importance of the lay press in the transmission of medical knowledge to the scientific community. N Engl J Med. 1991;325:1180–1183. http://dx.doi.org/10.1056/NEJM199110173251620
17. Casini S, Nerisini F. Behind closed doors: Scientists' and science communicators' discourses on science in society. A study across European research institutions. Technoscienza. 2012;3(2):37–62.
18. Mandavilli A. Trial by Twitter. Nature. 2011;469:268–269. http://dx.doi.org/10.1038/469286a
19. Marcinkowski F, Kohring M. The changing rationale of science communication : A challenge to scientific autonomy. JCOM. 2014;13(03), C04, 8 pages.
20. Massarani L, De Castro Moreira I. Science communication in Brazil: A historical review and considerations about the current situation. An Acad Bras Cienc. 2016;88(3):1577–1595. http://dx.doi.org/10.1590/0001-3765201620150338
21. Pandor N. South African science diplomacy: Fostering global partnerships and advancing the African agenda. Sci Diplomacy. 2012;1(1), 6 pages. http://www.sciencediplomacy.org/files/south_african_science_diplomacy.pdf
22. Dubow S. A commonwealth of knowledge: Science, sensibility and white South Africa 1820 – 2000. New York: Oxford University Press; 2006.
23. Sooryamoorthy R. Science and scientific collaboration in South Africa: Apartheid and after. Scientometrics. 2010;84(2):373–390. http://dx.doi.org/10.1007/s11192-009-0106-y
24. Barnard CN. The operation. A human cardiac transplant: An interim report of a successful operation performed at Groote Schuur Hospital, Cape Town. S Afr Med J. 1967;41(48):1271–1274.
25. Brink JG, Hassoulas J. The first human heart transplant and further advances in cardiac transplantation at Groote Schuur Hospital and the University of Cape Town. Cardiovasc J Afr. 2009;20(1):31–35.
26. South African Department of Science and Technology (DST). White paper on science and technology. Pretoria: DST; 1996.
27. Philander SG. How many scientists does South Africa need? S Afr J Sci. 2009;105(5/6):172–173.
28. South African Department of Science and Technology (DST). Science engagement framework. Pretoria: DST; 2014.
29. House of Lords. Science and society. London: House of Lords; 2000.
30. Poliakoff E, Webb TL. What factors predict scientists' intentions to participate in public engagement of science activities? Sci Commun. 2007;29(2):242–263. https://doi.org/10.1177/1075547007308009
31. Johnson DR, Ecklund EH, Lincoln AE. Narratives of science outreach in elite contexts of academic science. Sci Commun. 2014;36(1):81–105. https://doi.org/10.1177/1075547013499142
32. Reddy V, Gastrow M, Juan A, Roberts B. Public attitudes to science in South Africa. S Afr J Sci. 2013;109(1/2), Art. #1200, 8 pages. http://dx.doi.org/10.1590/sajs.2013/1200

33. Claassen G. Science and the media in South Africa: Reflecting a 'dirty mirror.' Communicatio. 2011;37(3):351–366. https://doi.org/10.1080/02500167.2011.622288

34. Gething L. 'Them and us': Scientists and the media – attitudes and experiences. S Afr Med J. 2003;93(3):197–201.

35. Bauer M, Jensen P. The mobilization of scientists for public engagement. Public Underst Sci. 2011;20(1):3–11. https://doi.org/10.1177/0963662510394457

36. Crettaz von Roten F. Gender differences in scientists' public outreach and engagement activities. Sci Commun. 2011;33(1):52–75. https://doi.org/10.1177/1075547010378658

37. Kreimer P, Levin L, Jensen P. Popularization by Argentine researchers: The activities and motivations of CONICET scientists. Public Underst Sci. 2010;20(1):37–47. https://doi.org/10.1177/0963662510383924

38. Chikoore L, Probets S, Fry J, Creaser C. How are UK academics engaging the public with their research? A cross-disciplinary perspective. High Educ Q. 2016;70(2):145–169. https://doi.org/10.1111/hequ.12088

39. Marcinkowski F, Kohring M, Fürst S, Friedrichsmeier A. Organizational influence on scientists' efforts to go public. Sci Commun. 2014;36(1):56–80. http://dx.doi.org/10.1177/1075547013494022

40. South African Department of Science and Technology (DST). South African National Survey of Research and Experimental Development – main analysis report 2014/15. Pretoria: DST; 2015.

41. Statistics South Africa (StatsSA). Mid-year population estimates. Pretoria: StatsSA; 2016.

42. Ceci SJ, Williams WM. Understanding current causes of women's underrepresentation in science. Proc Natl Acad Sci USA. 2011;108(8):3157–3162. https://doi.org/10.1073/pnas.1014871108

43. Butler-Adam J. Africa needs more women hooked on science [homepage on the Internet]. c2015 [cited 2017 Jan 17]. Available from: http://mg.co.za/article/2015-06-04-africa-needs-to-get-more-women-hooked-on-science

44. The Royal Society. Factors affecting science communication. London: The Royal Society; 2006.

45. Crettaz von Roten F, Goastellec G. Understanding academics' popular science publishing: Institution culture and management style effects. JNGS. 2015;13(2):15–29.

46. Dunwoody S, Brossard D, Dudo A. Socialization or rewards? Predicting US scientist-media interactions. Journalism Mass Comm. 2009;86(2):299–314. https://doi.org/10.1177/107769900908600203

47. Entradas M, Bauer M. Mobilisation for public engagement: Benchmarking the practices of research institutes. Public Underst Sci. 2016;26(7):771–788. https://doi.org/10.1177/0963662516633834

48. Neresini F, Bucchi M. Which indicators for the new public engagement activities? An exploratory study of European research institutions. Public Underst Sci. 2011;20(1):64–79. https://doi.org/10.1177/0963662510388363

49. Times Higher Education. World university rankings 2016 [homepage on the Internet]. c2016 [cited 2017 Mar 17]. Available from: https://www.timeshighereducation.com/world-university-rankings/2016/world-ranking

50. Peters H, Spangenberg A, Lo YY. Variations of scientist-journalist interactions across academic fields: Results of a survey of 1600 German researchers from the humanities, social sciences and hard sciences. In: Bucchi M, Trench B, editors. Quality, honesty and beauty in science and technology communication. Florence: PCST Network; 2012. p. 257–263.

51. Maze K, Barnett M, Botts EA, Stephens A, Freedman M, Guenther L. Making the case for biodiversity in South Africa : Re-framing biodiversity communications. Bothalia. 2016;46(1), Art. #2039, 8 pages. https://doi.org/10.4102/abc.v46i1.2039

52. Clark R, Partridge T, Kuman K. Caves of the ape-men: South Africa's Cradle of Humankind World Heritage Site. Johannesburg: Wits University Press; 2010.

53. Shisana O, Rehle T, Simbayi LC, Zuma K, Jooste S, Zungu N, et al. South African national HIV prevalence, incidence and behaviour survey. Pretoria: HSRC Press; 2014.

54. Peters HP. Scientists as public experts: Expectations and responsibilities. In: Bucchi M, Trench B, editors. Routledge handbook of public communication of science and technology. New York: Routledge; 2014. p. 70–82.

55. Berger LR, Hawks J, De Ruiter DJ, Churchill SE, Schmid P, Delezene LK, et al. *Homo naledi*, a new species of the genus *Homo* from the Dinaledi Chamber, South Africa. eLife. 2015;10(4), e09560, 35 pages. https://doi.org/10.7554/eLife.09560

56. Lents NH. Paleoanthropology wars. Skeptic Magazine. 2016;21(2):8–11.

57. McKie R. Scientist who found new human species accused of playing fast and loose with the truth [homepage on the Internet]. c2015 [cited 2017 Jan 05]. Available from: https://www.theguardian.com/science/2015/oct/25/discovery-human-species-accused-of-rushing-errors

58. Foster G. *Homo naledi*: piltdown and a lesson in African prehistory [homepage on the Internet]. c2015 [cited 2017 Jan 23]. Available from: http://thoughtleader.co.za/gavinfoster/2015/11/19/naledi-piltown-and-a-lesson-in-african-prehistory/

59. Noakes TD. Low-carbohydrate and high-fat intake can manage obesity and associated conditions: Occasional survey. S Afr Med J. 2013;103(11):826–830. https://doi.org/10.7196/SAMJ.7302

60. Naude CE, Schoonees A, Senekal M, Young T, Garner P, Volmink J. Low carbohydrate versus isoenergetic balanced diets for reducing weight and cardiovascular risk: A systematic review and meta-analysis. PLoS ONE. 2014;9(7), e100652, 30 pages. https://doi.org/10.1371/journal.pone.0100652

61. Sboros M. Tim Noakes – backlash as UCT academics say he's a danger to the public [homepage on the Internet]. c2014 [cited 2016 Dec 14]. Available from: http://www.biznews.com/health/2014/08/25/tim-noakes-backlash-uct-academics-say-hes-danger-public/

62. Wild S. Enter the rockstar scientist, exit trust in science [homepage on the Internet]. c2015 [cited 2017 May 03]. Available from: https://africacheck.org/2015/10/14/comment-enter-the-rockstar-scientist-exit-peoples-trust-in-science/

63. Eliseev A. We need more scientific rock stars, not fewer [homepage on the Internet]. c2015 [cited 2017 May 07]. Available from: https://www.dailymaverick.co.za/article/2015-10-18-op-ed-we-need-more-scientific-rock-stars-not-less/

64. Hassen T. Professor Noakes' name cleared [homepage on the Internet]. c2017 [cited 2017 May 15]. Available from: http://www.iol.co.za/southern-suburbs-tatler/news/professor-noakes-name-cleared-8837443

South Africa's geothermal energy hotspots inferred from subsurface temperature and geology

AUTHORS:
Taufeeq Dhansay[1,2]
Chiedza Musekiwa[1]
Thakane Ntholi[1]
Luc Chevallier[1]
Doug Cole[1]
Maarten J. de Wit[2]

AFFILIATIONS:
[1]Council for Geoscience, Bellville, South Africa
[2]Africa Earth Observatory Network – Earth Science Stewardship Research Institute, Nelson Mandela University, Port Elizabeth, South Africa

CORRESPONDENCE TO:
Taufeeq Dhansay

EMAIL:
taufeeq.dhansay@gmail.com

DATES:
Received: 24 Mar. 2017
Revised: 22 May 2017
Accepted: 07 July 2017

KEYWORDS:
geothermometry; tectonics; renewable energy policy

HOW TO CITE:
Dhansay T, Musekiwa C, Ntholi T, Chevallier L, Cole D, De Wit MJ. South Africa's geothermal energy hotspots inferred from subsurface temperature and geology. S Afr J Sci. 2017;113(11/12), Art. #2017-0092, 7 pages. http://dx.doi.org/10.17159/sajs.2017/20170092

ARTICLE INCLUDES:
✓ Supplementary material
× Data set

FUNDING:
None

South Africa intends to mitigate its carbon emissions by developing renewable energy from solar, wind and hydro, and investigating alternative energy sources such as natural gas and nuclear. Low-enthalpy geothermal energy is becoming increasingly popular around the world, largely as a result of technological advances that have enabled energy to be harnessed from relatively low temperature sources. However, geothermal energy does not form part of South Africa's future renewable energy scenario. This omission may be related to insufficient regional analysis of potentially viable geothermal zones across the country. We considered existing subsurface temperature and heat flow measurements and performed solute-based hydrochemical geothermometry to determine potentially anomalous geothermal gradients that could signify underlying low-enthalpy geothermal energy resources. We correlated these findings against hydro/geological and tectonic controls to find prospective target regions for investigating geothermal energy development. Our results show a significant link between tectonic features, including those on-craton, and the development of geothermal potential regions. In addition, potential regions in South Africa share similarities with other locations that have successfully harnessed low-enthalpy geothermal energy. South Africa may therefore have a realistic chance of developing geothermal energy, but will still need additional research and development, including new temperature measurements, and structural, hydrogeological and economic investigations.

Significance:

- The regional low-enthalpy geothermal energy potential of South Africa should be further researched for consideration of low-enthalpy geothermal energy as a renewable energy option.

Introduction

South Africa is the leading carbon emitter in Africa and has one of the highest rates of emissions of nations in the world.[1] This status can be linked to South Africa's vast coal resources, which are an important contributor to the local mining sector and also account for more than 80% of South Africa's energy generation.[1] South Africa intends to reduce its carbon emissions by producing about 40% of the country's total energy through renewable sources by 2030.[1] This goal will be achieved mostly through solar-, wind- and hydro-generated forms of energy and largely accelerated by a Renewable Energy Independent Power Producer Procurement Programme, which has attracted considerable private-sector investment.[1] Renewable energy alone will not meet South Africa's growing energy demands and therefore the country will also consider additional large-scale coal-fired energy, nuclear energy and energy produced from shale gas.[1]

Low-enthalpy geothermal energy is becoming increasingly popular around the world.[2] This popularity is largely because it requires geothermal gradients as low as ca 40 °C/km, which may be found in many global settings. South Africa does not have any active or recent volcanism and is situated far from any active continental and/or oceanic plate boundaries, but does have anomalously high heat flow regions that could meet the requirements for low-enthalpy geothermal energy development.[3-5]

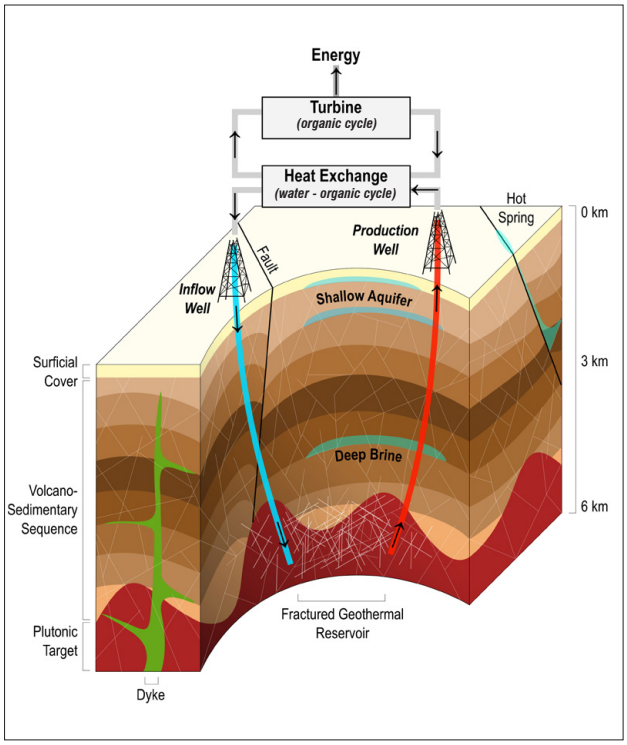
In this study, we aimed to elaborate on potentially viable geothermal regions of South Africa. To do this, we considered existing heat flow, heat productivity, downhole temperature and hot spring data to conduct estimates of the geothermal gradient across South Africa. We calculated the geothermal gradient using thermodynamic principles for historical heat flow and heat productivity data and from solute-based geothermometry on hot spring hydrochemical data. We also correlated these results with high heat producing plutonic and volcano-sedimentary rocks, and established underlying tectonic influences using regional seismicity. We used these results to present a geothermal potential map of South Africa and we made recommendations toward including low-enthalpy geothermal energy in South Africa's future renewable energy mix scenario.

Low-enthalpy geothermal energy

Geothermal resources can be broadly classified into convective and conductive systems. These systems describe regions of the upper crust that exhibit anomalously high heat flow, and either have naturally occurring and/or circulating groundwater (i.e. convective), or are typically dry (i.e. conductive). Low-enthalpy geothermal resources represent systems in which groundwater circulating from a reservoir would not reach the surface with a temperature above ca 100 °C. High-enthalpy geothermal resources, on the other hand, are generally limited to global locations with active plate tectonics and consequentially active/recent volcanism, and where groundwater is heated to near and above supercritical levels. Low-enthalpy resources are usually associated with ancient tectonic activity and are often defined by plutonic rocks with high concentrations of heat-producing radiogenic elements (e.g. uranium and potassium) which are overlain by a thick and insulating volcano/sedimentary sequence. These conditions are commonly found in most parts of the world and may account for the increase in global low-enthalpy geothermal exploration (for more details refer to Huenges and Ledru[2]).

Harvesting heat from a low-enthalpy geothermal resource commonly uses a binary mechanism with two independent and separated working fluids. In general, a geothermal fluid is circulated through a porous fractured reservoir in a targeted high heat producing plutonic assembly. Here simultaneous sequestration is also possible, for example through the incorporation of carbon dioxide in the geothermal fluid. Once the geothermal fluid is adequately heated, it is brought to the surface where it enters a generation plant. Within the generation plant, the heated geothermal fluid enters a heat-exchange mechanism under pressure and interacts with a second organic fluid that has a much lower boiling point. Conductive heat transfer causes the secondary fluid to flash to steam, which is then used to produce energy. Thereafter, the organic condensate is returned to the heat exchange system while the cooled geothermal fluid is cycled back into the fractured reservoir (Figure 1).

A comparative example of low-enthalpy geothermal energy development that may be considered here is within the Upper Rhine Graben (URG) along the border between Germany and France. The URG highlights extension along the Alpine foreland and in the Landau geothermal region (southwest Germany); it consists of fractured Palaeozoic basement granite with uranium content of up to ca 10 PPM[6], overlain by ca 1.5-km thick Cenozoic, Mesozoic and Permian sedimentary rock sequences[7]. Rifting makes the URG seismically active[8], with significant crustal thinning that enables uplift linked to mantle upwelling[9]. The average geothermal gradient throughout the URG is ca 35–45 °C/km with high heat flow evident from numerous hot springs. Hydrogeological properties throughout the URG are highly complicated[10]; however, the average groundwater yield rate as measured around geothermal sites and at a depth of ca 2 km is approximately 0.1 L/s.[11] An average 5 MW low-enthalpy geothermal plant in the URG produces from reservoir temperatures of about 130 °C at an average depth of 3.5–4.5 km and production flow rates of 40–130 L/s.[11] Heated water is typically used to run an Organic Rankine Cycle generation system with excess hot water diverted to provide household heating. There is approximately 30 MW of installed low-enthalpy geothermal capacity within the URG with exploration and development continuing to increase.[11]



Source: Modified after Dhansay et al.[3]; refer to Huenges and Ledru[2] for more details.

Figure 1: Schematic illustration of a binary fluid enhanced geothermal system related to surrounding fracture-controlled geological features.

Geological controls on heat flow in South Africa

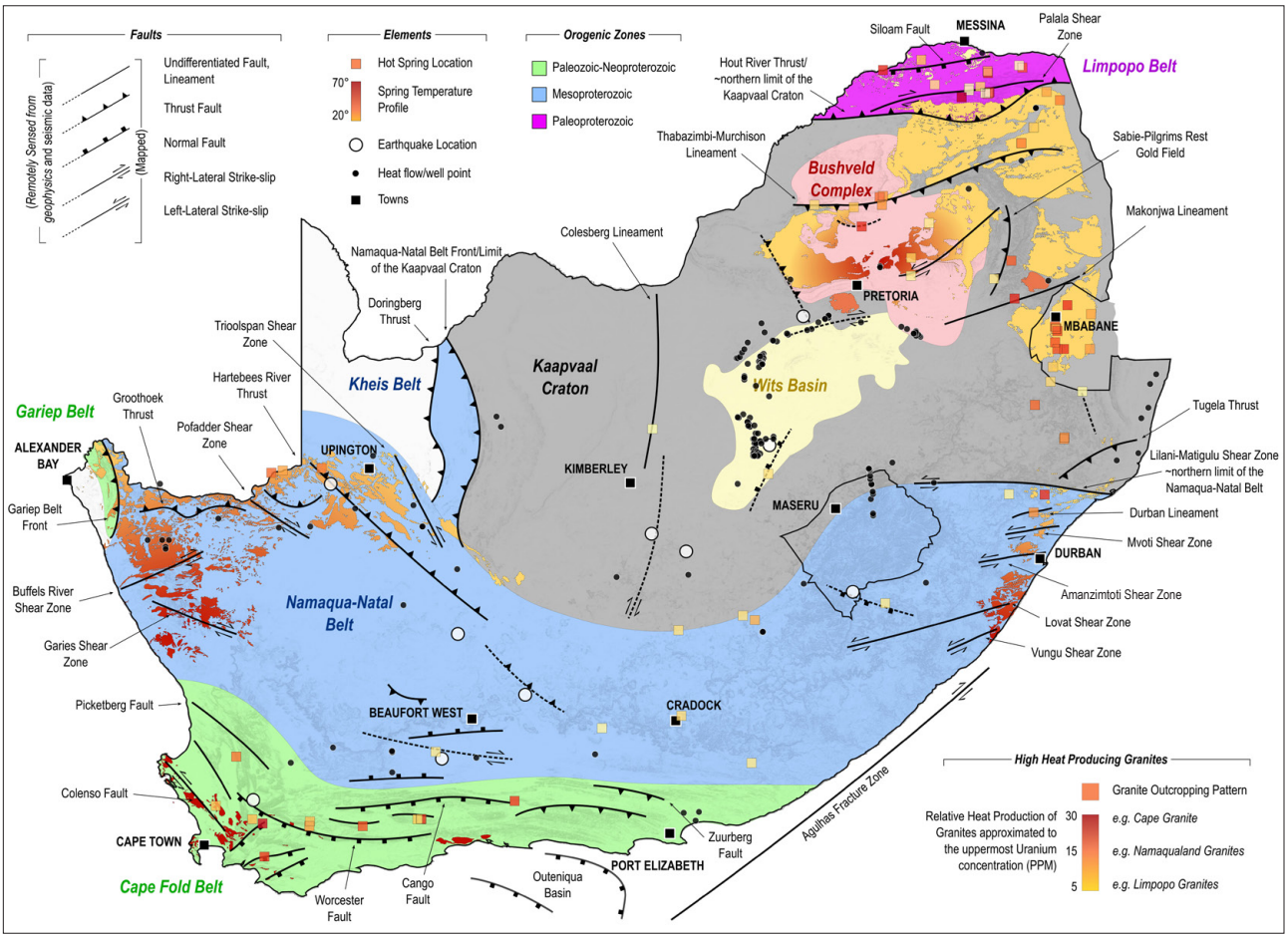
South Africa is partially underlain by the Kaapvaal Craton and its thick subcontinental lithospheric mantle keel that reaches depths of up to ca 250 km and has an average crustal thickness of 40–50 km[12] (Figure 2). The Kaapvaal Craton comprises several smaller fragments of ancient crust that amalgamated and stabilised during the early Archean. Regions where amalgamation occurred appear as deep crustal discontinuities that may be likened to more recent plate tectonic boundaries.[13] In general, the Kaapvaal Craton has a relatively low heat flow[14-16], which has largely discouraged extensive geothermal investigation. However, subsurface temperature data suggest that there is at least some evidence for low-enthalpy geothermal energy potential on the Kaapvaal Craton[4], and especially on the surrounding palaeo-orogenic belts[3]. These orogenic belts demarcate regions where continental collision had occurred. Regions showing apparent low-enthalpy geothermal energy potential, especially orogenic belts, share several characteristic geological and tectonic similarities. Most notably, orogenic belts display significantly higher heat flow signatures.[3] This characteristic is especially illustrated by a ca 60 mW/m^2 heat flow increase across the boundary of the Kaapvaal Craton and the Namaqua-Natal Belt[17,18], and similarly across the Limpopo Belt[3].

Each orogenic belt is associated with tectonic evolutionary processes related to different supercontinent cycles; for example, the Limpopo Belt formed during the amalgamation of the Kalahari Craton[13]; the Namaqua-Natal and Gariep Belts formed during the formation of Rodinia[19]; and the Cape Fold Belt formed during the formation of Gondwana[20]. During these events convergent-related subduction resulted in the emplacement of partial melt-derived plutonic rocks, many of which are rich in heat-producing elements that release heat during the decay of radiogenic elements (Figure 2). For example, the Cape Granite Suite (Cape Fold Belt) has uranium concentrations of up to ca 34 PPM[21]; the Namaqua-Natal Belt has uranium concentrations of ca 10–54 PPM[22,23]; even older Archean granite-gneisses around Mombela (Nelspruit)[24] and Johannesburg[25] exhibit uranium concentrations of up to ca 20–28 PPM. In addition, Palaeoproterozoic tectonic activity along the Thabazimbi-Murchison Lineament[26] may have assisted in the emplacement of the Bushveld Complex, which includes felsic rocks that exhibit uranium concentrations of up to 30 PPM[27].

Post-convergent extensive forces resulted in the formation of volcano-sedimentary basins that overlie and insulate radiogenic plutonic rocks, and often exhibit their own elevated heat-producing signatures, particularly related to elevated and economically significant uranium concentrations, e.g. the Karoo Basin (largely overlying the Cape Fold Belt and the Namaqua-Natal Belt)[28]; the Soutpansberg (overlying the Limpopo Belt) and Springbok Flats (overlying the Bushveld)[28]. Significantly elevated radiogenic signatures are also evident within the on-craton Archean Witwatersrand and Pongola Basin strata[29]; and especially from the Palaeoproterozoic Transvaal rocks. Here, partial melt derived products associated with the emplacement of the Bushveld Complex sometimes highlight anomalous uranium concentrations of up to 250 PPM.[27]

Higher heat flow signatures are further corroborated by numerous hot springs concentrated along orogenic belts and below the Karoo escarpment (Figure 2). Heating and circulation of groundwater is enabled by complex brittle fracture networks that were formed and reactivated during various plate tectonic events, e.g. hot springs located in the Limpopo Belt use a fracture network that was largely created during the Palaeoproterozoic and which underwent reactivation several times, including during more recent Mesozoic uplift.[30]

Another important link between the orogenic zones is an elevated number of natural seismic events[31] which highlight stress release-reactivation along deep-seated brittle structures (Figure 2). For example, seismicity in the Karoo may be correlated with structures in the underlying Namaqua-Natal basement[32]; and anomalous radon release within the Cape Fold Belt[33], in addition to the occurrence of historically significant seismic events[34], highlight the influence of stress release along deep structures.



Sources: Tectonic information largely derived from 19, 26, 35–38

Figure 2: Overview of the major tectonic structures and zones across South Africa with the locations of significant earthquake focal mechanisms and inferred structures related to these events. Locations of the various data sources used within this study (e.g. hot springs and temperature measurement points) and high heat producing plutonic rocks are also highlighted. Note that the Namaqua-Natal Belt probably continues beneath the Cape Fold Belt as far as the offshore Agulhas Fracture Zone.[12]

Geothermal gradient calculations

In this study, we considered available data (Figure 2; also refer to supplementary material), including heat flow, heat productivity, downhole temperature measurements and solute-based equilibria geothermometry to highlight prospective zones for investigating low-enthalpy geothermal energy development. We normalised across the various data sources by calculating the theoretical geothermal gradient and using inverse distance weighting to interpolate these results across South Africa. For hot springs with only surface temperature information, we estimated circulation depths of ca 2–5 km, which we inferred from shallow geophysical investigations[39,40], surrounding heat flow measurements and from deep borehole temperature data[41]. We also identified important tectonic structures and estimated the (most recent) underlying faulting dynamics based on earthquake focal mechanisms.

We calculated the geothermal gradient from heat flow and heat productivity data using principles of thermodynamics, where Q represents the heat flow (mW/m^2) and C represents the lithological thermal conductivity (mW/m/°C). Where no thermal conductivity measurements were available, we made estimations based on experimental thermal conductivity calculations[42] and the known geological profiles. These factors are related by Fourier's Law:

$$\frac{dT}{dZ} = \frac{Q}{C} \qquad \text{Equation 1}$$

We also calculated the geothermal gradient by applying solute-based equilibria geothermometry on available hot spring hydrochemical data and relate these results against the inferred hot spring circulation depths. Solute-based geothermometry estimates hot spring reservoir temperatures using the presence of equilibrated mineral cations, particularly silica, sodium and potassium.[43] Importantly, unknown fluid-rock interactions and/or sporadic infiltration/flow rates insinuate that hot spring reservoirs are not likely to be in a state of equilibrium and therefore this technique may not provide exact reservoir temperatures. Nevertheless, this method is still useful to establish a general estimate of hot spring reservoir temperatures.[43] We used geothermometry limited to a maximum allowable temperature of 250 °C, including the silica-cation geothermometer:

$$T = \left(\frac{1309}{5.19 - logSi}\right) - 273.15 \qquad \text{Equation 2}$$

where T is the reservoir temperature and Si is the concentration of dissolved silica in the water. We also used the Na-K geothermometer for springs with insignificant silica content and/or if silica content was not measured, where Na and K represent the concentration of dissolved sodium and potassium, respectively:

$$T = \left(\frac{1217}{\log\left(\frac{Na}{K}\right) + 1.483}\right) \qquad \text{Equation 3}$$

Anomalous heat flow regions in South Africa

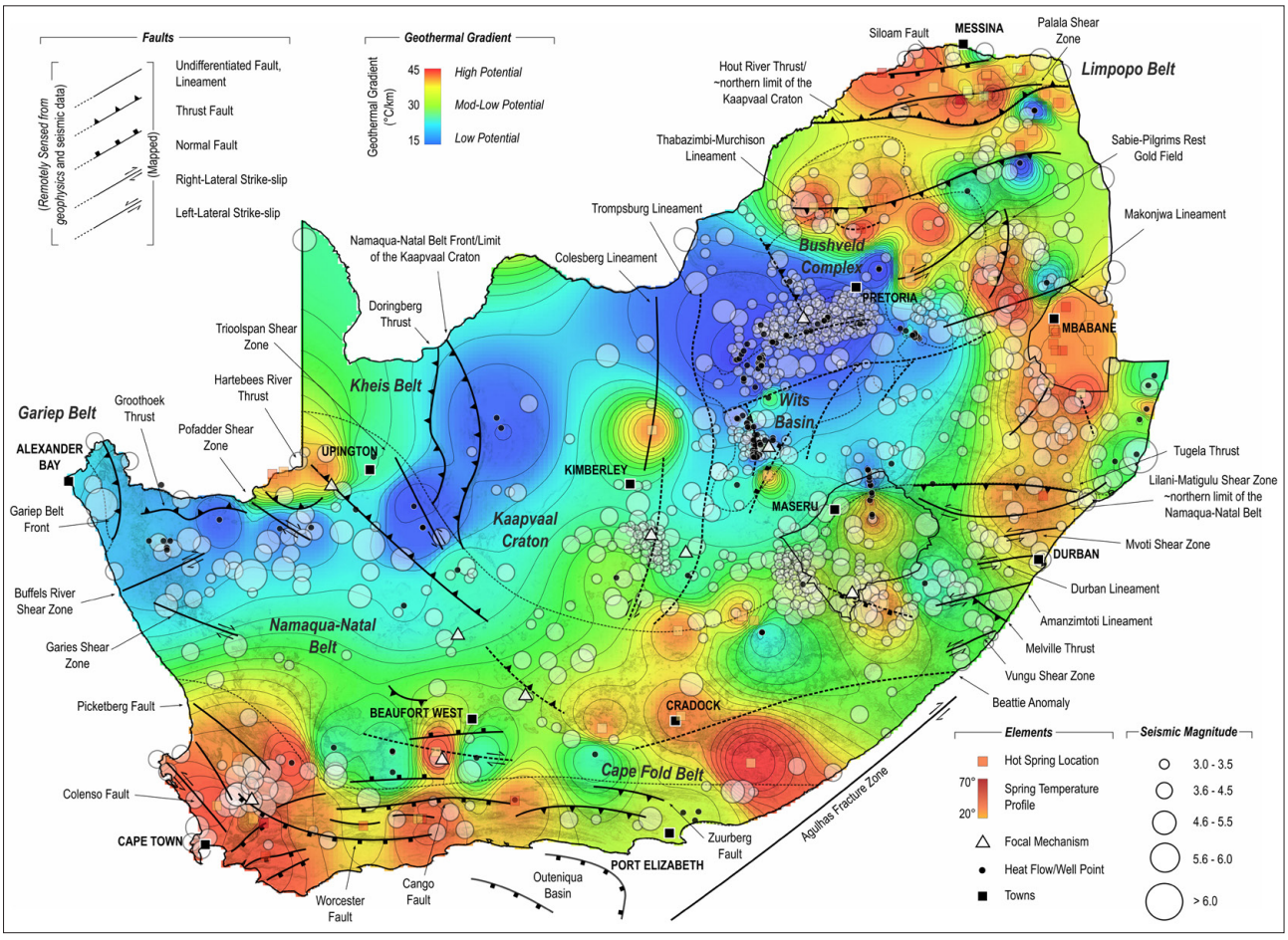
The results of the geothermal gradient calculations are summarised in Figure 3. In general, the highest calculated geothermal gradients are closely related to naturally occurring seismic events and are situated within orogenic belts surrounding the Kaapvaal Craton. In addition, these orogenic zones account for the largest number of hot springs in South Africa. Anomalous geothermal gradients and hot springs are also found on the Kaapvaal Craton and are notably related to mapped cratonic discontinuities (e.g. the Colesberg and Thabazimbi-Murchison Lineaments). There is also a strong correlation between high geothermal gradients and the outcrop pattern of highly radiogenic plutonic rocks, particularly where these are overlain by volcano-sedimentary sequences (Figure 4).

Discussion

In general, orogenic belts surrounding the Kaapvaal Craton exhibit the highest heat flow signatures in South Africa, which may be linked to underlying geological, tectonic and crustal compositional controls, particularly related to the production of radiogenic material. These orogenic belts experience varying phases of convergent and extensive tectonics that often result in: the emplacement of high heat producing plutonic rocks[21-23]; the development of sedimentary basins, which were infilled by mostly siliciclastic sediments interspersed with often highly radiogenic volcanic extrusive material[28]; and finally the creation of complex brittle structural networks that enable thermal convective dispersion through natural groundwater flow[50] and seismic-inducing stress release[32].

Our results also highlight that high heat flow is not only restricted to off-craton regions. Zones near deep cratonic discontinuities also exhibit elevated geothermal gradients, which is especially noticeable along the Colesberg, Thabazimbi-Murchison and Makonjwa Lineaments. These zones highlight more ancient tectonic activity associated with the amalgamation and stabilisation of the Kaapvaal Craton[13] – processes that have apparently also imparted elevated heat flow signatures. These on-craton regions also have high heat producing plutonic rocks (e.g. Archean granite-gneisses around Mombela and Johannesburg; and Palaeoproterozoic Bushveld felsic rocks) and overlying volcano-sedimentary basin sequences that exhibit their own high heat producing signatures (e.g. the Witwatersrand, Pongola, Transvaal and Springbok Flats). Anomalous heat flow and underpinning tectonic influences are also evident with the occurrence of hot springs and higher number of natural on-craton seismic events around these structures (Figure 4d). In general, seismicity related to these deep cratonic discontinuities highlight reactivation associated with a more recent northeast to southwest oriented extension[31], which is in agreement with the present-day stress state seen in much of South Africa[32].

Using the results of this study, together with high-yielding, shallow groundwater aquifers as a proxy for deeper hydrogeological conditions, and considering factors of successful development in Germany, we may highlight the most promising regions for investigating low-enthalpy geothermal energy development in South Africa (Figure 4). In no particular order, these areas especially include, but are not necessarily restricted to: (1) regions of the Cape Mountains, especially the Syntaxial region; (2) the southern Karoo; (3) the boundary of the Namaqua-Natal Belt and Kaapvaal Craton north of Durban; (4) the Bushveld Basin near the Thabazimbi-Murchison Lineament, north of Johannesburg; (5) the Limpopo Belt.



Sources: Temperature data derived from 5, 14–18, 44–59 and references therein.

Figure 3: Graphical overview of the calculated geothermal gradients across South Africa. Map includes major tectonic contacts and structures, seismic activity and earthquake focal mechanisms and hot spring locations.

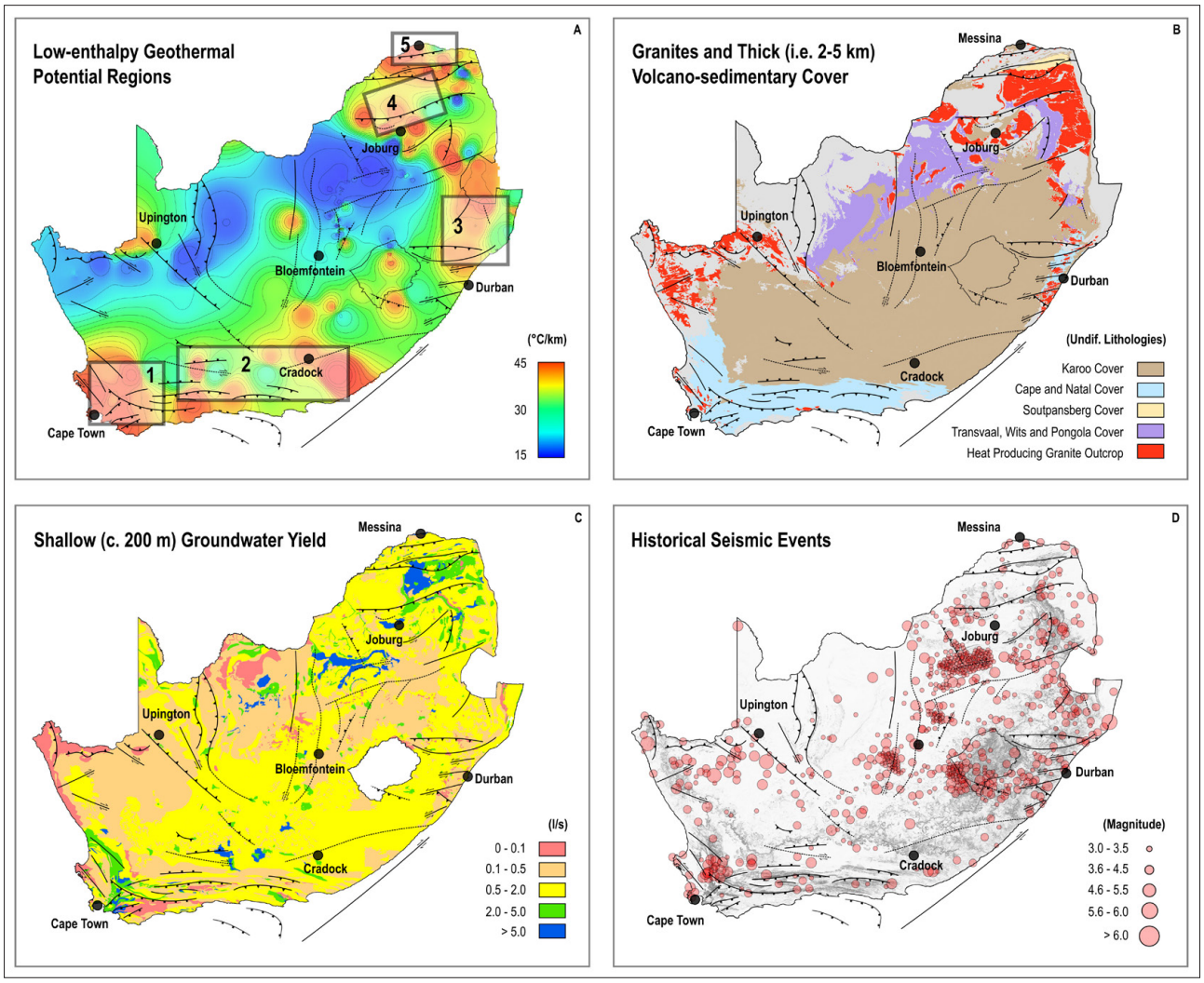


Figure 4: (a) Potentially viable low-enthalpy geothermal investigation regions (1–5); based on (b) high heat producing plutonic rocks and overlying volcano-sedimentary rocks; and (c) approximate groundwater yield (data from the South African Department of Water Affairs). (d) Regional seismicity (data from the Council for Geoscience).

Importantly, the cost of initial exploration and development of low-enthalpy geothermal energy is high[3] and development in Germany was largely enabled by a Governmental Renewable Energy Feed-In Tariff of 15 EURc/kWh[60]. The impact of financial incentives in South Africa is also noticeable with the Renewable Energy Independent Power Producer Procurement Programme, which has resulted in the cost of wind and solar being reduced by 46% and 71%, respectively.[1] Including geothermal in this programme could potentially accelerate further research and development and may result in geothermal being added to South Africa's future energy mix.

Conclusions and recommendations

The results of this study suggest that despite geothermal (re)sources not being part of South Africa's renewable energy mix, the country does have some potential for harnessing low-enthalpy geothermal energy. We therefore recommend that South Africa seriously considers geothermal energy as another renewable option. However, there are several key factors that need to be addressed before harvesting of geothermal energy can occur.

South Africa still needs significant research and data acquisition, including: high-resolution ground-based geophysics, new and extensive downhole temperature measurements, structural mapping, and deep hydrogeological and isotope hydrochemical investigations. These data will allow for a more precise evaluation of South Africa's geothermal energy potential and also highlight any possible negative impacts, especially on groundwater quality and inducing seismicity.[8] Finally, economic modelling is imperative to design mechanisms to adequately enable advanced geothermal research and development in South Africa.

Acknowledgements

We acknowledge the support provided by the Department of Science and Technology and Council for Geoscience (South Africa), and the Jena Structure and Tectonics Research Group (Germany). We appreciate the Reviewers' comments and suggestions that greatly improved this manuscript. This article is an Iphakade and AEON contribution, numbers 166 and 168, respectively.

Authors' contributions

T.D. was the lead researcher and primary author; M.D. provided supervision and reviewed the manuscript; C.M., T.N., L.C. and D.C. reviewed the final manuscript.

References

1. South African Department of Energy. Annual report 2015/16. Pretoria: Department of Energy; 2016. Available from: http://www.energy.gov.za/files/Annual%20Reports/DoE-Annual-Report-2015-16.pdf

2. Huenges E, Ledru P. Geothermal energy systems: Exploration, development, and utilization. Hoboken, NJ: John Wiley & Sons; 2011. http://dx.doi.org/10.1002/9783527630479.

3. Dhansay T, De Wit M, Patt A. An evaluation for harnessing low-enthalpy geothermal energy in the Limpopo Province, South Africa. S Afr J Sci. 2014;110(3/4), Art. #2013-0282, 10 pages. http://dx.doi.org/10.1590/sajs.2014/20130282.

4. Ntholi T. A technical and economic evaluation of a Passive Underground Mine-water Purification System (PUMPS): A geothermally powered geo-engineering system designed for in-situ bio-remediation of acid mine water [PhD thesis]. Port Elizabeth: Nelson Mandela Metropolitan University; 2017.

5. Campbell SA, Mielke P, Götz AE. Geothermal energy from the Main Karoo Basin? New insights from borehole KWV-1 (Eastern Cape, South Africa). Geoth Energy. 2016;4(1), Art. #9, 19 pages. http://dx.doi.org/10.1186/s40517-016-0051-y.

6. Hooijkaas GR, Genter A, Dezayes C. Deep-seated geology of the granite intrusions at the Soultz EGS site based on data from 5km-deep boreholes. Geothermics. 2006;35(5):484–506. http://dx.doi.org/10.1016/j.geothermics.2006.03.003.

7. Sissingh W. Comparative tertiary stratigraphy of the Rhine Graben, Bresse Graben and Molasse Basin: Correlation of Alpine foreland events. Tectonophysics. 1998;300(1):249–284. http://dx.doi.org/10.1016/s0040-1951(98)00243-1.

8. Grünthal G. Induced seismicity related to geothermal projects versus natural tectonic earthquakes and other types of induced seismic events in Central Europe. Geothermics. 2014;52:22–35. http://dx.doi.org/10.1016/j.geothermics.2013.09.009.

9. Freymark J, Sippel J, Scheck-Wenderoth M, Bär K, Stiller M, Fritsche JG, Kracht M. The deep thermal field of the Upper Rhine Graben. Tectonophysics. 2017;694:114–129. http://dx.doi.org/10.1016/j.tecto.2016.11.013.

10. Stober I, Bucher K. Hydraulic and hydrochemical properties of deep sedimentary reservoirs of the Upper Rhine Graben, Europe. Geofluids. 2015;15(3):464–482. http://dx.doi.org/10.1111/gfl.12122

11. Agemar T, Weber J, Schulz R. Deep geothermal energy production in Germany. Energies. 2014;7(7):4397–4416. http://dx.doi.org/10.3390/en7074397.

12. Stankiewicz J, De Wit M. 3.5 billion years of reshaped Moho, southern Africa. Tectonophysics. 2013;609:675–689. http://dx.doi.org/10.1016/j.tecto.2013.08.033.

13. Schmitz MD, Bowring SA, De Wit MJ, Gartz V. Subduction and terrane collision stabilize the western Kaapvaal craton tectosphere 2.9 billion years ago. Earth Planet Sci Lett. 2004;222(2):363-376. http://dx.doi.org/10.1016/s0012-821x(04)00220-1.

14. Krige LJ. Borehole temperatures in the Transvaal and Orange Free State. Proc Roy Soc Lond A. 1939;173(955):450–474. http://dx.doi.org/10.1098/rspa.1939.0158.

15. Carte AE, Van Rooyen AI. Further measurements of heat flow in South Africa. Geol Soc S Afr. Special Publication. 1969;2:445–448.

16. Jones MQ. Heat flow in the Witwatersrand Basin and environs and its significance for the South African shield geotherm and lithosphere thickness. J Geophys Res-Sol Ea. 1988;93(B4):3243–3260. http://dx.doi.org/10.1029/jb093ib04p03243.

17. Jones MQ. Heat flow and heat production in the Namaqua mobile belt, South Africa. J Geophys Res-Sol Ea. 1987;92(B7):6273–6289. http://dx.doi.org/10.1029/jb092ib07p06273.

18. Jones MQ. Heat flow anomaly in Lesotho: Implications for the southern boundary of the Kaapvaal craton. Geophys Res Lett. 1992;19(20):2031–2034. http://dx.doi.org/10.1029/92gl02207.

19. Thomas RJ. A tale of two tectonic terranes. S Afr J Geol. 1989;92(4):306–321.

20. Linol B, De Wit MJ. Origin and evolution of the Cape Mountains and Karoo Basin. Regional Geology Reviews series. Cham, Switzerland: Springer International Publishing; 2016. http://dx.doi.org/10.1007/978-3-319-40859-0.

21. Viljoen J, Siegfried H. Kuils River Batholith. In: Johnson M, editors. Pretoria: Catalogue of South African Lithostratigraphic Units; 2009. p. 13.

22. Voordouw RJ, Rajesh HM. Granitoids from the Margate Terrane and their implications for tectono-magmatic models of the Natal Metamorphic Province (South Africa). S Afr J Geol. 2012;115(1):47–64. http://dx.doi.org/10.2113/gssajg.115.1.47.

23. Andreoli MA, Hart RJ, Ashwal LD, Coetzee H. Correlations between U, Th content and metamorphic grade in the western Namaqualand Belt, South Africa, with implications for radioactive heating of the crust. J Petrol. 2006;47(6):1095–1118. http://dx.doi.org/10.1093/petrology/egl004.

24. Meyer M, Robb LJ, Anhaeusser CR. Uranium and thorium contents of Archaean granitoids from the Barberton Mountain Land, South Africa. Precambrian Res. 1986;33(4):303–321. http://dx.doi.org/10.1016/0301-9268(86)90048-3.

25. Van Tonder DM, Mouri H. Petrology and geochemistry of the granitoid rocks of the Johannesburg Dome, central Kaapvaal Craton, South Africa. S Afr J Geol. 2010;113(3):257–286. http://dx.doi.org/10.2113/gssajg.113.3.257.

26. Good N, De Wit MJ. The Thabazimbi-Murchison Lineament of the Kaapvaal craton, South Africa: 2700 Ma of episodic deformation. J Geol Soc London. 1997;154(1):93–97. http://dx.doi.org/10.1144/gsjgs.154.1.0093.

27. Walraven F, Kleeman GJ, Allsopp HL. Disturbance of trace-element and isotope systems and its bearing on mineralisation in acid rocks of the Bushveld Complex, South Africa. High heat production (HHP) granites, hydrothermal circulation and ore genesis. London: Institute of Mining and Metallurgy; 1985. p. 393–408.

28. Kenan AO, Chirenje E. Uranium in South Africa: Exploration and supply capacity. Nat Res Conserv. 2016;4(2):25–33. http://www.hrpub.org/journals/article_info.php?aid=5013.

29. Robb LJ, Davis DW, Kamo SL, Meyer FM. Ages of altered granites adjoining the Witwatersrand Basin with implications for the origin of gold and uranium. Nature. 1992;357(6380):677–680. http://dx.doi.org/10.1038/357677a0.

30. Dhansay T, Brandl G, De Wit MJ. Fractal geometry of the fault network across the Soutpansberg Mountains, Limpopo, South Africa. S Afr J Geol. 2016;119(1):235–242. http://dx.doi.org/10.2113/gssajg.119.1.235.

31. Brandt MB, Saunders I. New regional moment tensors in South Africa. Seismol Res Lett. 2011;82(1):69–80. http://dx.doi.org/10.1785/gssrl.82.1.69.

32. Viola G, Kounov A, Andreoli MA, Mattila J. Brittle tectonic evolution along the western margin of South Africa: More than 500Myr of continued reactivation. Tectonophysics. 2012;514:93–114. http://dx.doi.org/10.1016/j.tecto.2011.10.009.

33. Nemangwele F. Radon in the Cango Caves [MSc thesis]. Cape Town: University of the Western Cape; 2005. http://etd.uwc.ac.za/xmlui/handle/11394/231.

34. Goedhart ML, Booth PW. A palaeoseismic trench investigation of early Holocene neotectonic faulting along the Kango Fault, southern Cape Fold Belt, South Africa – Part I: Stratigraphic and structural features. S Afr J Geol. 2016;119(3):545–568. http://dx.doi.org/10.2113/gssajg.119.3.545.

35. Dingle RV, Scrutton RA. Continental breakup and the development of post-Paleozoic sedimentary basins around southern Africa. Geol Soc Am Bull. 1974;85(9):1467–1474. http://doi.org/10.1130/0016-7606(1974)85<1467:cbatdo>2.0.co;2.

36. Boshoff R, Reenen DV, Smit CA, Perchuk LL, Kramers JD, Armstrong R. Geologic history of the Central Zone of the Limpopo Complex: The West Alldays area. J Geol. 2006;114(6):699–716. http://dx.doi.org/10.1086/507615.

37. Eglington BM. Evolution of the Namaqua-Natal Belt, southern Africa – A geochronological and isotope geochemical review. J Afr Earth Sci. 2006;46(1):93–111. http://dx.doi.org/10.1016/j.jafrearsci.2006.01.014.

38. Voordouw RJ. A D3 shear zone in the Margate Terrane and its implications for regional deformation in the Natal Metamorphic Province (South Africa). S Afr J Geol. 2010;113(2):183–194. http://dx.doi.org/10.2113/gssajg.113.2.183.

39. Baiyegunhi C, Mupandawana NW, Oloniniyi TL, Gwavava O. Characterization of Aliwal North Hotspring in the Eastern Cape Province of South Africa, using magnetic, electromagnetic and radiometric methods. IOSR J Eng. 2014;4(12):43–58. http://dx.doi.org/10.9790/3021-041203043058.

40. Nyabeze PK, Gwavava O. Investigating heat and magnetic source depths in the Soutpansberg Basin, South Africa: Exploring the Soutpansberg Basin Geothermal Field. Geoth Energy. 2016;4(1):1–20. http://dx.doi.org/10.1186/s40517-016-0050-z.

41. Verhagen B, Butler M, Levin M, Van Wyk E. Environmental isotopes assist in groundwater sustainability assessment of the Taaibosch fault zone, Northern Province, South Africa. In: Sililo O, editor. Groundwater: Past achievements and future challenges. Rotterdam: Balkema; 2000. p. 673–678.

42. Clauser C, Huenges E. Thermal conductivity of rocks and minerals. In: Ahrens T, editor. Rock physics & phase relations: A handbook of physical constants. Washington DC: American Geophysical Union; 1995. p. 105–126. http://dx.doi.org/10.1029/rf003.

43. Ellis AJ. Chemical geothermometry in geothermal systems. Chem Geol. 1979;25(3):219–226. http://dx.doi.org/10.1016/0009-2541(79)90143-8.

44. Bullard EC. Heat flow in South Africa. R Soc Lond Proc Ser A Math Phys Eng Sci. 1939:474–502.

45. Kent LE. The thermal waters of the Union of South Africa and South West Africa. S Afr J Geol. 1949;52(1):231–264.

46. Carte AE. Heat flow in the Transvaal and Orange Free State. P Phys Soc Lond B. 1954;67(9):664. http://dx.doi.org/10.1088/0370-1301/67/9/302.

47. Bouwer R. Borehole temperatures in the Klerksdorp and Orange Free State areas. Bulletin 22. Pretoria: Geological Survey of the Union of South Africa Department of Mines; 1954. p. 35.

48. Gough DI. Heat flow in the southern Karroo. Proc Roy Soc Lond A. 1963;272(1349):207–230. http://dx.doi.org/10.1098/rspa.1963.0050.

49. Hoole J. The development of Lilani hot springs: An analysis of socio-economic and environmental impacts [MSc thesis]. Pietermaritzburg: University of Natal; 2000. p. 165. http://researchspace.ukzn.ac.za/xmlui/handle/10413/3799.

50. Woodford A, Chevallier L. Hydrogeology of the Main Karoo Basin: Current knowledge and future research needs. WRC report no. TT179/02. Pretoria: Water Research Commission; 2002. p. 482.

51. Merkel B, Steinbruch F. Preliminary hydrogeological investigation on the Nhambita hot spring, Mozambique. Report to the Department of Scientific Services, Gorongosa National Park. 2007. p. 33. Available from: http://www.gorongosa.org/our-story/science/reports/preliminary-hydrogeological-investigation-nhambita-hot-spring-mozambique.

52. Grootjans AP, Grundling PL, Grundling A, Linström A, Engelbrecht J, Price JS. Spring mires fed by hot artesian water in Kruger National Park, South Africa. Mires and peat. 2010;6(7):1–10. http://pixelrauschen.de/wbmp/media/map06/map_06_07.pdf.

53. Viljoen J, Stapelberg F, Cloete M. Technical report on the geological storage of carbon dioxide in South Africa. Pretoria: Council for Geoscience; 2010. p. 236.

54. Olivier J, Venter JS, Jonker CZ. Thermal and chemical characteristics of hot water springs in the northern part of the Limpopo Province, South Africa. Water SA. 2011;37(4):427–436. http://dx.doi.org/10.4314/wsa.v37i4.1.

55. Boekstein M. Revitalising the healing tradition – health tourism potential of thermal springs in the Western Cape [Doctor of Technology thesis]. Cape Town: Cape Peninsula University of Technology; 2012. p. 182. http://digitalknowledge.cput.ac.za/xmlui/handle/11189/1313.

56. Olivier J, Jonker CZ. Optimal utilisation of thermal springs in South Africa. WRC report no. TT577. Pretoria: Water Research Commission; 2013. p. 13.

57. Robins N. The thermal springs of Swaziland: A review. In: Groundwater: A New Paradigm. Proceedings of the Geological Society of South Africa Biennial Conference; 2013 September 17–19; Durban, South Africa. Johannesburg: Geological Society of South Africa; 2013. p. 7. http://nora.nerc.ac.uk/503375/.

58. Hicks N, Davids S, Beck B, Green A. Investigation of CO_2 storage potential of the Durban Basin in South Africa. Energy Procedia. 2014;63:5200–5210. http://dx.doi.org/10.1016/j.egypro.2014.11.551.

59. De Kock M, Beukes N, Götz A, Cole D, Robey K, Birch A, et al. Open file progress report on exploration of the Southern Karoo Basin through CIMERA-KARIN borehole KZF-1 in the Tankwa Karoo, Witzenberg (Ceres) district. Johannesburg: University of Johannesburg; 2015. p. 12.

60. Bakhtyar B, Fudholi A, Hassan K, Azam M, Lim CH, Chan NW, et al. Review of CO_2 price in Europe using feed-in tariff rates. Renew Sust Energ Rev. 2017;69:685–691. http://dx.doi.org/10.1016/j.rser.2016.11.146.

Finding fossils in Malapa breccia – medical CT scanning or micro-CT scanning?

AUTHOR:
Jacqueline S. Smilg[1,2,3]

AFFILIATIONS:
[1]Department of Radiation Sciences, University of the Witwatersrand, Johannesburg, South Africa
[2]Evolutionary Studies Institute, School of Geosciences, University of the Witwatersrand, Johannesburg, South Africa
[3]Department of Radiology, Charlotte Maxeke Johannesburg Academic Hospital, Johannesburg, South Africa

CORRESPONDENCE TO:
Jacqueline Smilg

EMAIL:
jsmilg@yahoo.com

DATES:
Received: 22 Feb. 2017
Revised: 19 Apr. 2017
Accepted: 10 July 2017

KEYWORDS:
Australopithecus sediba; manual preparation; kVp; beam energy; image quality

HOW TO CITE:
Smilg JS. Finding fossils in Malapa breccia – medical CT scanning or micro-CT scanning? S Afr J Sci. 2017;113(11/12), Art. #2017-0057, 6 pages. http://dx.doi.org/10.17159/sajs.2017/20170057

ARTICLE INCLUDES:
✓ Supplementary material
× Data set

FUNDING:
Malapa Project, Evolutionary Studies Institute, University of the Witwatersrand (grant no. 98839)

Computed tomography (CT) imaging of fossils has revolutionised the field of palaeontology, allowing researchers to gain a better understanding of fossil anatomy, preservation and conservation. Micro focus X-ray computed tomography (μXCT) has been far more extensively used for these purposes than medical CT (XCT) – mostly because of the exquisite detail that the μXCT scanning modality, using slices of micron thicknesses, can produce. High energy X-rays can potentially penetrate breccia more effectively than lower energy beams. This study demonstrates that lower energy beams produce superior images for prioritising breccia for preparation. Additionally, XCT scanners are numerous, accessible, fast and relatively cost-effective when compared to μXCT scanners – the latter are not freely available, scanning times are much longer and there are significant limitations on the size and weight of scannable objects. Breccia blocks from Malapa were scanned at high and lower energy and images were analysed for image quality, artifact and certainty of diagnosis. Results show that lower energy images are deemed superior to higher energy images for this particular application. This finding, taken together with the limitations associated with the use of μXCT for the imaging of the large breccia from Malapa, shows that XCT is the better modality for this specific application. The ability to choose fossil-bearing breccia, ahead of manual mechanical preparation by laboratory technicians, would allow for the optimal use of limited resources, manual preparatory skills as well as the curtailment of costs.

Significance:

- 'Blind' manual preparation of fossil-bearing breccia is a costly and time-consuming exercise – and often results in a low yield.
- The ability to triage fossil-bearing breccia ahead of manual preparation would allow for the optimal use of limited resources.
- Medical CT is better than micro-CT to triage breccia to allow for prioritisation of rocks for manual preparation.

Background

The use of computed tomography (CT) in the analysis of fossils has become common place[1-4], although most of the CT work to date has been performed on prepared or partially prepared specimens[3,5,6]. The application of CT for matrix that potentially contains fossils has lagged behind these many advances in the visualisation and study of prepared fossils.

Micro CT (μXCT) scanners have become increasingly popular in the imaging of prepared fossils as a result of the combination of their ability to produce high-resolution images because of much smaller slice thicknesses (in the micron range), increased spatial resolution and more variable energy capabilities (especially in the higher energy ranges) than medical CT (XCT) scanners; in fact, μXCT dominates the current fossil literature as the X-ray modality of choice for virtual analysis of prepared fossils.[7-10]

The site of Malapa has yielded hundreds of breccia blocks, which have the potential to contain fossils of the hominin *Australopithecus sediba*. Traditionally, breccia has undergone manual preparation – with a single block taking months to adequately prepare, with no guarantee of obtaining any fossils. This process is costly – both in time and money – and is not a prudent use of scarce preparatory skills.

New ways have been sought to better deal with the many Malapa breccia requiring assessment. The idea of triage (prioritising for preparation) of breccia by imaging prior to manual preparation has been explored. Mass use of XCT to image Malapa breccia has been undertaken and the predicted findings from imaging have been compared to the actual post-preparation findings.[11] Smilg and Berger[11] demonstrated good correlation and suggested that XCT is a valuable imaging modality in the triage process. There is considerable advantage to being able to know the contents of a rock ahead of costly, time-consuming 'blind' manual preparation. Knowledge of breccia contents before manual preparation allows decisions to be made as to the most efficient use of resources, personnel and funds, in addition to allowing planning of the course of action desired for the preparation of fossil material.

The effectiveness of XCT with regard to image quality and object penetration was explored. The quality of an X-ray beam is the measurement of the penetrating power of the photons, which depends on the energy of the photons, the atomic number (Z), the density and the thickness of the object being scanned.[12] Peak kilovoltage (kVp) governs the penetrating power of photons – the higher the kVp, the more the beam penetrates the object. Hence the question was posed as to whether higher kilovoltage beams would produce images from the breccia that would be better quality than those produced from lower kilovoltage beams. XCT is limited in its kilovoltage, with the maximum being 140 kV, whereas μXCT is capable of higher kilovoltage. As a consequence of the better penetrative ability of high energy X-rays,[13] it may be expected that high energy scanning would be superior to lower energy scanning when it is applied to large, dense breccia blocks.

High and lower energy images of breccia were obtained to address this question of image quality from the Malapa breccia for triage purposes. μXCT was used to obtain the higher kilovoltage values that XCT was unable to generate. The spectrum of potential imaging characteristics available from μXCT was not researched for this particular application.

The image quality of lower energy and high energy images when applied to objects within breccia blocks was compared. Material for this experiment was selected from the fossil hominin bearing site of Malapa in the Cradle of Humankind World Heritage site because of the importance of fossils from this site, the many blocks collected from which there are no fossils visible on the surface, and the sheer number of blocks retrieved that require assessment.

The possibility of the use of neutron microtomography (NCT or n-μCT) or magnetic resonance imaging (MRI) was considered. NCT differs from X-rays in that neutrons can penetrate materials that are opaque to X-rays and organic material strongly attenuates these neutron beams. NCT may thus be appropriate for imaging organically preserved fossils as a complement to XCT or μXCT.[14] NCT has shown promising results in being able to differentiate otherwise similar dense materials – a recent study has documented the use of NCT to view a fossil encased in breccia.[15]

However, NCT may induce hazardous levels of radioactivity in some geological materials which leaves these imaged samples radioactive and necessitates that the samples be isolated for a long time after the imaging. Additionally, there are currently no functioning NCT machines in South Africa, where this breccia triage is needed. NCT necessitates much longer scanning times when compared to XCT and is also limited to smaller fields of view.

MRI maps properties related to the chemical environment of certain elements, rather than mapping radiation attenuation. MRI has been considered to be poorly suited to geological material[16] and at present does not compete with μXCT or NCT. MRI machines are not easily accessible to palaeontological researchers, scans are very expensive and scan times are long compared with those of XCT. For these reasons, both NCT and MRI were not considered suitable for the mass screening of breccia from Malapa.

Fossils may have the potential to contain ancient DNA and whilst the effect of radiation on living tissue has been well investigated, little has been done to research the impact that radiation may have on ancient DNA.[17] Recent work has shown that radiation of fossils may have a detrimental effect on ancient DNA when the total surface dose exceeds 200 Gray, so these researchers recommended using as low a dose as possible when scanning fossils as well as using resolution no higher than necessary to achieve the desired outcome.[18] The value of 200 Gray is far higher than any dose from a XCT or μXCT scan (8000 times higher than the highest dose for a medical CT scan).[18]

The Malapa site

The site of Malapa (site UW88)[19] represents an unusually rich early hominin locality in Africa[20-24], dating to 1.977 ± 0.002 million years ago (Ma)[25]. The site contains remains of several individuals, all ascribed to *A. sediba.*[20] These remains are found alongside an abundant, well-preserved fauna.[24] It is postulated that this well-preserved material was accumulated during a seemingly rapid depositional event that occurred over a few days, weeks or months.[26] The site of Malapa is located in the region of the Cradle of Humankind World Heritage Site, northwest of Johannesburg, South Africa. The locality is recognised as a de-roofed cave of at least 25 x 20 m, in an area where limited limestone mining had taken place, probably during the late 19th or early 20th century, almost certainly before Robert Broom began exploring the area in the mid-1930s.[26]

Materials and methods

A total of 15 breccia blocks from the site of Malapa were chosen to be scanned at differing kVp using both XCT and μXCT. Blocks were selected taking cognisance of the limitations on weight and volume presented by the University of the Witwatersrand's micro-CT scanner (a maximum permissible weight of 50 kg).

Lower energy scanning was performed on a Philips Brilliance 128 slice CT at Charlotte Maxeke Johannesburg Academic Hospital (South Africa), whilst high energy scanning was performed on a Nikon Metrology XTH 225/320 LC dual source industrial CT system at the University of the Witwatersrand. Both data sets were analysed on AMIRA 5.4.5 Ink.

This experiment was intended to simulate a real-world situation. After discussions with the principal scientists in charge of the Malapa project, it was decided that if mass screening of breccia was to be implemented, it would be impractical to adjust energy levels and scanning parameters for every block in order to optimise image quality, because of the desire to scan large numbers of rocks in a single session and the possibility that the scanning would be overseen by technicians not familiar with CT images. Hence the imaging parameters would have to be constant and pre-decided. The parameters selected are given in Table 1. A single energy parameter was selected for each system based upon initial tests that resulted in a good quality image from both machines. Reference was also made to prior work done with low energy scanning.[11] Micro-CT images were reconstructed with micro-CT Pro v2.2 associated with the Nikon Metrology XTH 225/320 LC dual source industrial CT system.

Table 1: Parameters used for scanning on medical CT (XCT) and micro-CT (μXCT)

XCT	μXCT
140 kVp	300 kVp
360 mAs	340 μA
1-mm slice thickness	4 frames per second
Automatic artifact correction	Ring artifact correction
Pitch 0.45	Frame averaging of 2
512 x 512 matrix	3000 projections
	Copper filtration of 3.4 mm

kVp, peak kilovoltage; mAs, milliamp seconds; µA, micro amps

Three objects, thought to represent potential fossils, were chosen from within each block for evaluation. The same object was identified on images from each modality and displayed on a two-dimensional image with the same orientation and alignment specific to the object under assessment. The viewing parameters (including greyscale setting and magnification) were fixed for all readers, with no manipulation allowed. Readers were supplied with identical static two-dimensional images. Qualitative visual assessment was done by each reader independently.

Each object was evaluated for overall image quality, certainty of diagnostic accuracy and imaging artifact. There is no objective definition of image quality – it is a matter of the observer's subjective judgement. CT artifacts can affect the quality of the images, sometimes to the point of making them diagnostically unusable. Artifact is any distortion or error in the image that is unrelated to the subject being studied.[12] Artifact production can degrade the CT image and hinder interpretation, but modern CT machines are developed with built-in artifact reduction features, including filters, calibration correction, automatic tube current modulation and scanner software.[27]

The images were rated according to a qualitative, visual five-point scale for each parameter given in Table 2.

The objective was to compare these parameters between lower kilovoltage and higher kilovoltage images. The scores from an individual reader for each object, per criterion, for each energy were compared, rather than analysing inter-reader variability. Results were evaluated by two diagnostic radiologists and three palaeontological scientists. These results are summarised in Appendices 1 and 2 in the supplementary material.

Table 2: Scale used to determine rating of image for each parameter assessed

Rating	Overall image quality	Diagnostic certainty	Artifact
5	Excellent delineation: 100%	Full and confident certainty	Absence of artifact
4	Clear/good delineation: >75–<100%	Good certainty	Mild artifact, not interfering with diagnosis
3	Adequate delineation: >50–<75%	Adequate certainty	Moderate artifact, slightly interfering with diagnosis
2	Marginally acceptable delineation: >25–<50%	Marginal certainty	Pronounced artifact, interfering with diagnosis but still possible to arrive at diagnosis
1	Unacceptable delineation: <25%	Uncertain	Artifact completely hinders diagnosis

Table 3: Summary of reader analysis showing number of reads for which readers scored the 140 kVp or 300 kVp image higher and the number of reads for which the images were scored equally. Overall, lower energy was deemed to produce better images when images were assessed for image quality, certainty of diagnosis and presence of artifact.

kVp scores	Artifact assessment reads	Certainty assessment reads	Quality assessment reads	Fraction of total reads	Percentage for kVp setting
140 kVp scored higher	114	136	151	401/630	63.65
300 kVp scored higher	27	14	25	66/630	10.48
140 kVp score = 300 kVp score	84	30	49	163/630	25.87

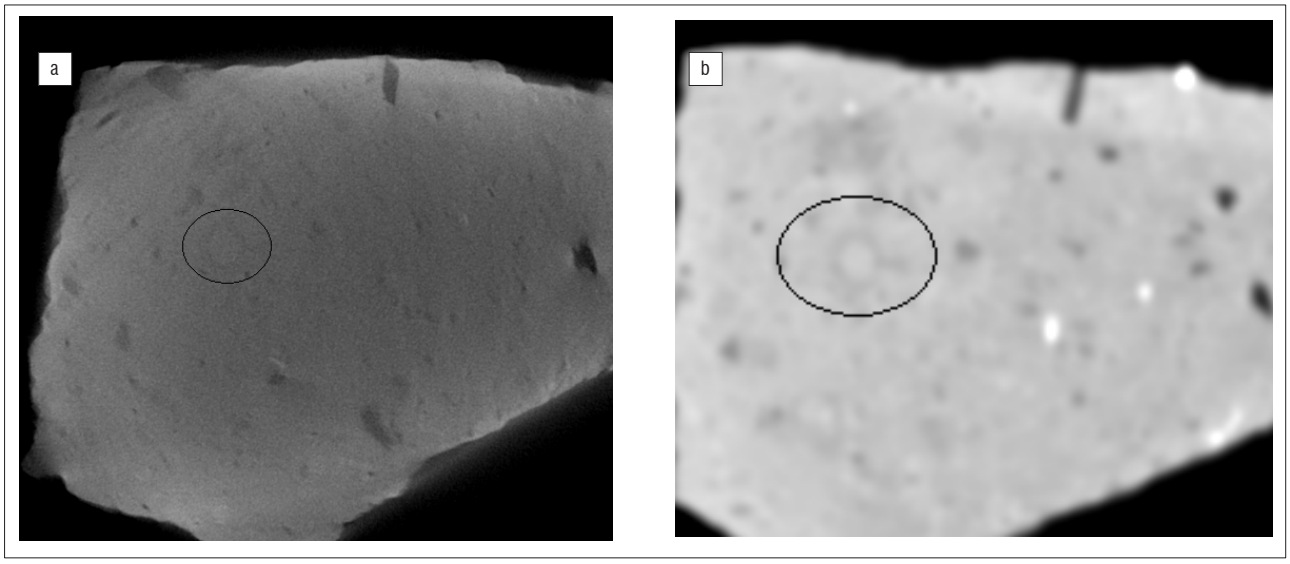


Figure 1: Object 2 within block 3195 (ringed). The potential fossil is visualised in the same plane and alignment on both modalities: (a) 300 kVp and (b) 140 kVp. Overall the readers deemed the image from lower energy to be better with fewer artifacts than that from higher energy.

Results

Results were assessed to determine which kVp choice scored higher (indicating better overall image) per reader, per object and per criterion, and whether the reader assigned the same score per criterion per object – the latter indicating an indeterminate result of neither kilovoltage setting producing a superior image (Table 3).

When considering the presence or absence of artifacts, the readers found that the lower energy scans produced better interpretive images in 114 of 225 reads (Figure 1). In 27 of 225 reads, high energy scanning was found to produce better interpretive images and 84 of 225 reads were indeterminate as to one scanning energy being better than the other.

Overall quality was deemed to be better in 151 of 225 reads for lower energy scans, whilst higher energy scans were deemed better in 25 of 225 reads (Figure 2). An indeterminate result, with both energy levels scoring equally, was obtained in 49 of 225 reads.

One reader did not rate for certainty of diagnosis for either modality because of a lack of confidence in cross-sectional identification. Results here identified lower energy scans as being superior in 136 of 180 reads (Figure 3). Higher energy scans were deemed better in 14 of 180 reads and results were indeterminate in 30 of 180 reads.

Overall percentages indicate that for all three criteria, the images obtained from lower energy scanning produced better, more diagnostic and more useful images with fewer artifacts in 63.65% of reads. Higher energy was deemed to deliver better images in 10.48% of reads and both modalities yielded equitable images in 25.87% of reads.

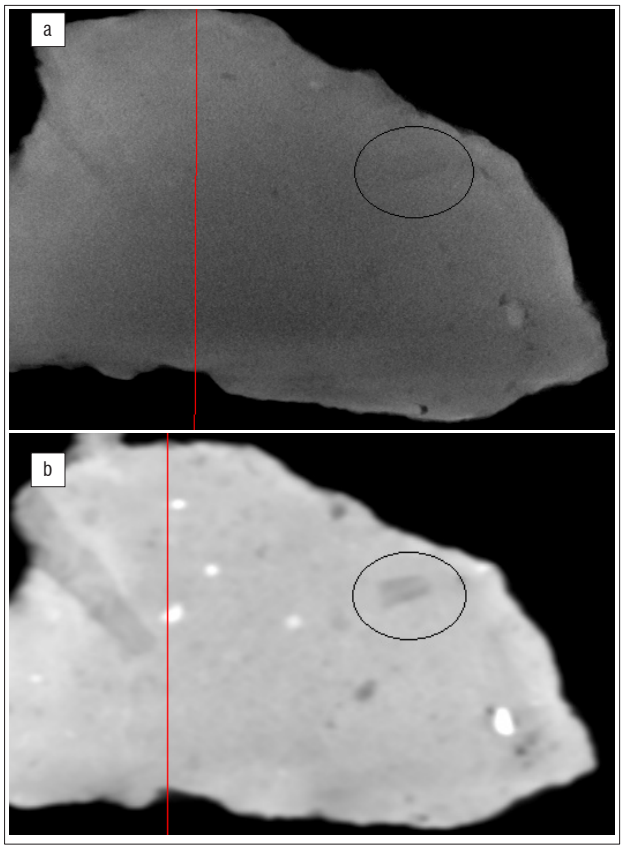


Figure 2: Object 3 within block 1343 (ringed). The potential fossil is visualised in the same plane and alignment on both modalities: (a) 300 kVp and (b) 140 kVp. Overall the readers deemed the image from lower energy to have better image quality than that from higher energy.

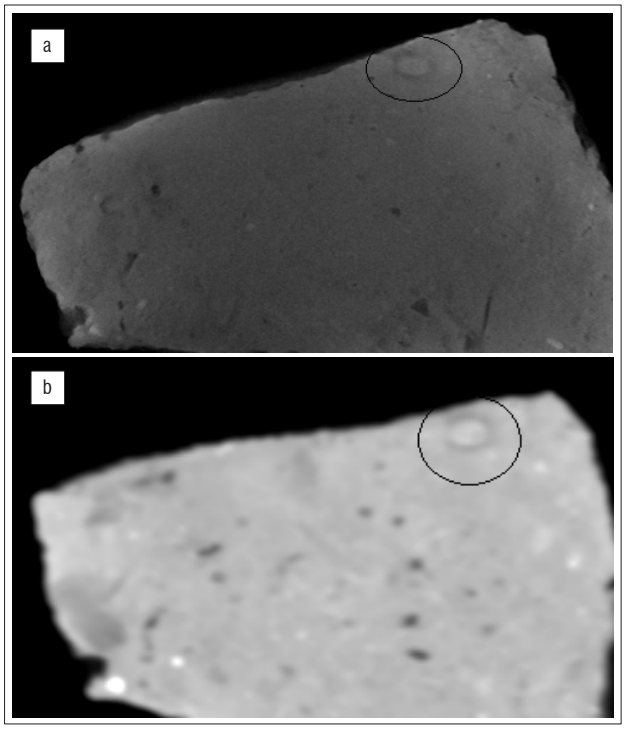


Figure 3: Object 2 within block 3728 (ringed). The potential fossil is visualised in the same plane and alignment on both modalities: (a) 300 kVp and (b) 140 kVp. Overall the readers deemed the image from lower energy to offer more certainty of diagnosis than that from higher energy.

Discussion

In producing a CT image, both milliamp seconds (mAs) and kVp must be chosen to ensure sufficient delivery of X-rays to achieve acceptable image quality.[28] The selection of kVp sets the energy of the X-rays reaching the object. The energy of an X-ray defines its penetrative ability[13], as well as the expected attenuation of the ray as it passes through materials of different densities. High energy X-rays, when compared to lower energy rays, may penetrate certain objects more effectively but are less sensitive to changes in material density and composition.

Each object varies in density and atomic number – both of which impact the X-ray beams' attenuation. Lower energy settings decrease the overall signal and result in increased noise, which can degrade the image. Increasing technical factors, such as mAs or kVp, decrease image noise but also increase dose. Because of the inanimate and inorganic nature of the material being studied, dose was not as important as it would be with living material. As X-rays traverse an object they are attenuated by scattering and absorption. This attenuation is a result of three processes[13]: photoelectric absorption, Compton scattering and pair production. Medical scanners (which use lower X-ray energy up to 140 kV) have more photoelectric effect and lower Compton scatter, whereas in higher energy μXCTs, attenuation is greater from Compton scatter.[13] This difference is important as it explains why low energy X-rays are more sensitive to differences in composition than higher energy ones. The photoelectric absorption is proportional to Z^{4-5} (atomic number of atom in the material to the power 4-5) whereas Compton absorption is proportional only to Z (atomic number).[29] Thus two materials could be differentiated in lower energy CT, but at higher energies may be indistinguishable. Despite a better penetration ability, high energy scanning may not differentiate objects as well as lower energy scanning.

The quality of a CT image is affected by several factors; enhancing or suppressing any of these factors depends upon the imaging interests.[12] CT image quality is dependent upon balancing these parameters to produce the best possible image for the object being scanned. CT parameters can be manipulated to either decrease or eliminate the adverse effects of these characteristics. Generally, there is a trade-off when CT parameters are manipulated.

Although CT scanners using high kVp (μXCT) could potentially be more effective at penetrating large breccia and can produce images with increased resolution as a result of much thinner slices, their disadvantages are their limitations in the size/weight of scannable objects, the long scan times needed as well as their limited availability to many scientists. In addition, the huge databases generated from the scan data necessitate specialised computer hardware and software for analysis, which are costly and not readily available to many researchers.

In contrast, medical CT machines are accessible as a consequence of their widespread use in clinical medical situations and are found in most hospitals and radiology practices, have a weight restriction of up to 200 kg and offer fast, reliable imaging and greater throughput of scanned objects. The databases are manageable on basic modern computers and data can be assessed with easily available (often free) software packages. Finally, at present, medical CT scanning, on a case by case basis, is typically significantly cheaper than micro-CT scanning, largely because of differences in initial machine cost as well as the presence of a greater number of medical CT machines in the community.

To be effective for the screening of Malapa breccia, the imaging modality used should:

- be cost effective, both in time and money
- be quick
- be repeatable
- use predetermined parameters that would negate the need for specialist attendance at scanning sessions
- be regularly readily accessible
- be able to accommodate large pieces of breccia.

The μXCT machines in the Gauteng area that are available to the Malapa project are limited in that:

- they have prolonged application and waiting times
- they are costly
- scanning is time consuming ($\pm$6 h/block)
- there are significant limitations in the weight and dimensions of objects to be scanned
- they generate large databases which make assessment time consuming and post processing limited to specialised programmes.

Conclusions

Fossils, particularly of hominins, are highly sought-after objects in the search for human origins, but 'blind' manual mechanical preparation of fossil-bearing breccia is a costly and time-consuming exercise – often with a low yield. The ability to triage breccia ahead of manual preparation would allow for the optimal use of limited resources, manual preparatory skills as well as the curtailment of costs.

Microcomputed tomography is well established as an imaging modality for the imaging of prepared fossils. Because of its potential to penetrate breccia more effectively by using high energy X-rays, it might be considered the modality of choice for breccia triage.

This study shows that in the application of breccia triage, the penetrating ability of a lower energy beam is not detrimental to the outcome.

Given the many other limitations of μXCT faced by the Malapa team, it is of significant advantage to researchers interested in a high throughput of potentially low value material, in search of high value material (as is the case with using CT scanners to triage potentially fossil-bearing blocks), that the differential penetration of lower and higher energy beams does not have a significant impact on image quality and that XCT is overall the better choice over μXCT for this application. The study does not seek to generalise the contribution of XCT and acknowledges that for other applications, μXCT may be the modality of choice. But for the purpose of fossil identification within large rocks and for breccia triage for the breccia originating from the Malapa site, it has been shown that XCT is superior to μXCT for this particular palaeontological application. Application of these findings can now be expanded to breccia from other fossil sites.

Additionally, as the effects of radiation on ancient DNA are still not clear, consideration should be given to the accumulative dose of radiation to which an individual fossil is exposed. It is recommended that the lowest possible dose necessary to achieve the desired outcome is used, as well as the lowest resolution possible to achieve the desired result. The dose increases at about the square power of the increase of resolution.[18] Thus, another reason to advocate the application of lower energy scanning over higher energy scanning for breccia triage is that the potential and ability to access ancient DNA from fossil specimens still needs research and elucidation.

Use of medical CT scanning of fossil-bearing breccia is thus an alternative to random block preparation. In order to maximise the use of limited resources and manual preparatory skills, as well as to curtail costs, this research shows that, prior to manual preparation, blocks should undergo scanning with medical CT scanners and virtual assessment of contents should be undertaken to allow for prioritisation of rocks for manual preparation.

Triage of fossil-bearing breccia using medical CT scanners will shorten the time from breccia removal from the field to the retrieval of relevant fossil specimens for interrogation, publication and dissemination of their information.

Acknowledgements

L. Berger, B. de Klerk, K. Jakata, J. Mukanku, E. Odes, P. Randolph-Quinney, F. Thackeray and R. van der Merwe of the Evolutionary Studies Institute at the University of the Witwatersrand (Johannesburg, South Africa) and National Palaeosciences Centre of Excellence are acknowledged for their assistance. J. Haberfeld and Q. Letsoalo of the Department of Diagnostic Radiology at Charlotte Maxeke Johannesburg Academic Hospital (Johannesburg, South Africa) are acknowledged for their assistance. The Virtual Imaging in Palaeontology Laboratory, the Microfocus X-ray Computed Tomography Facility at the Evolutionary Studies Institute and Charlotte Maxeke Johannesburg Academic Hospital are acknowledged for providing access to their facilities.

References

1. Sutton MD, Rahman IA, Garwood RJ. Techniques for virtual palaeontology. London: John Wiley & Sons; 2013. https://doi.org/10.1002/9781118591192
2. Lautenschlager S. Reconstructing the past. Methods and techniques for the digital restoration of fossils. R Soc Open Sci. 2016;3(10), Art. #160342, 18 pages. https://doi.org/10.1098/rsos.160342
3. Balzeau A, Crevecoeur I, Rougier H, Froment A, Gilissen E, Grimaud-Herv D, et al. Applications of imaging methodologies to paleoanthropology: Beneficial results relating to the preservation, management and development of collections. CRPalevol. 2010;9(6):265–275. https://doi.org/10.1016/j.crpv.2010.07.006
4. Wu X, Schepartz LA. Application of computed tomography in palaeoanthropological research. Prog Nat Sci. 2009;19:913–921.https://doi.org/10.1016/j.pnsc.2008.10.009
5. Odes EJ, Randolph-Quinney PS, Steyn M, Throckmorton Z, Smilg JS, Zipfel B, et al. Earliest hominin cancer: 1.7-million-year-old osteosarcoma from Swartkrans Cave, South Africa. S Afr J Sci. 2016;112(7/8), Art. #2015-0471, 5 pages. https://doi.org/10.17159/sajs.2016/20150471
6. Randolph-Quinney PS, Williams SA, Steyn M, Meyer MR, Smilg JS, Churchill SE, et al. Osteogenic tumour in *Australopithecus sediba*: Earliest hominin evidence for neoplastic disease. S Afr J Sci. 2016;112(7/8), Art. #2015-0470, 7 pages. https://doi.org/10.17159/sajs.2016/20150470
7. Biolaurus. 8 Spectacular examples of micro-CT being used to analyse fossils [homepage on the Internet]. No date [updated 2017; cited 2017 Jan 25]. Available from: http://biolaurus.com/8-spectacular-examples-micro-ct-used-analyze-fossils/
8. Liu Y, Scholtz G, Hou X. When a 520 million-year-old Chengjiang fossil meets a modern micro-CT – a case study. Sci Rep. 2015;5, Art. #12802, 8 pages. https://doi.org/10.1038/srep12802
9. Cunningham JA, Rahman IA, Lautenschlager S, Rayfield EJ, Donoghue PJC. A virtual world of paleontology. Trends Ecol Evol. 2014;29(6):347–357. https://doi.org/10.1016/j.tree.2014.04.004
10. Abel RL, Laurini CR, Richter MA. A palaeobiologist's guide to 'virtual' micro-CT preparation. Palaeontol Electron. 2012;15(2):496–500.
11. Smilg JS, Berger LR. Discovering hominins – Application of medical computed tomography (CT) to fossil-bearing rocks from the site of Malapa, South Africa. PLoS ONE. 2015;10(12), e0145340, 19 pages. https://doi.org/10.1371/journal.pone.0145340.
12. Reddinger W. CT image quality [document on the Internet]. c1998 [cited 2017 Feb 22]. Available from: https://www.coursehero.com/file/9938632/CT-Image-Quality/
13. Ketcham RA, Carlson WD. Acquisition, optimization and interpretation of X-ray computed tomographic imaging: Application to the geosciences. Comput GeoSci. 2001;27(4):381–400. https://doi.org/10.1016/S0098-3004(00)00116-3
14. Winkler B. Applications of neutron radiography and neutron tomography. Rev Mineral Geochem. 2006;63:459–471. https://doi.org/10.2138/rmg.2006.63.17
15. Beaudet A, Braga J, De Beer F, Schillinger B, Steininger C, Vodopivec V, et al. Neutron microtomography-based virtual extraction and analysis of a cercopithecoid partial cranium (STS 1039) embedded in a breccia fragment from Sterkfontein member 4 (South Africa). Am J Phys Anthropol. 2016;159(4):737–745. https://doi.org/10.2/ajpa.22916
16. Sutton MD. Tomographic techniques for the study of exceptionally preserved fossils. Proc R Soc B Biol Sci. 2008;275(1643):1587–1593. https://doi.org/10.1098/rspb.2008.0263

17. Grieshaber BM, Osborne DL, Doubleday AF, Kaestle FA. A pilot study into the effects of X-ray and computed tomography exposure on the amplification of DNA from bone. J Archaeol Sci. 2008;35(3):681–687. https://doi.org/10.1016/j.jas.2007.06.001

18. Immel A, Le Cabec A, Bonazzi M, Herbig A, Temming H, Schuenemann VJ, et al. Effect of X-ray irradiation on ancient DNA in sub-fossil bones – Guidelines for safe X-ray imaging. Sci Rep. 2016;6, Art. #32969, 14 pages. https://doi.org/10.1038/srep32969

19. Zipfel B, Berger LR. New cenozoic fossil-bearing site abbreviations for collections in the University of the Witwatersrand. Palaeont Afr. 2009;44:77–81.

20. Berger LR, De Ruiter DJ, Churchill SE, Schmid P, Carlson KJ, Dirks PHGM, et al. *Australopithecus sediba*: A new species of *Homo*-like australopith from South Africa. Science. 2010;328(5975):195–204. https://doi.org/10.1126/science.1184944

21. Kibii JM, Churchill SE, Schmid P, Carlson KJ, Reed ND, De Ruiter DJ, et al. A new partial pelvis of *Australopithecus sediba*. Science. 2011;333(6048):1407–1411. https://doi.org/10.1126/science.1202521

22. Zipfel B, De Silva JM, Kidd RS, Carlson KJ, Churchill SE, Berger LR. The foot and ankle of *Australopithecus sediba*. Science. 2011;333(6048):1417–1420 https://doi.org/10.1126/science.1202703

23. Kivell TL, Kibii JM, Churchill SE, Schmid P, Berger LR. *Australopithecus sediba* hand demonstrates mosaic evolution of locomotor and manipulative abilities. Science. 2011;333(6048):1411–1417. https://doi.org/10.1126/science.1202625

24. Kuhn BF, Werdelin L, Hartstone-Rose A, Lacruz RS, Berger LR. Carnivoran remains from the Malapa hominin site, South Africa. PLoS ONE. 2011;6(11), e26940, 11 pages. https://doi.org/10.1371/journal.pone.0026940

25. Pickering R, Dirks PHGM, Jinnah Z, De Ruiter DJ, Churchill SE, Herries AIR, et al. *Australopithecus sediba* at 1.977 Ma and implications for the origins of the genus *Homo*. Science. 2011;333(6048):1421–1423. https://doi.org/10.1126/science.1203697

26. Dirks PHGM, Kibii JM, Kuhn BF, Steininger C, Churchill SE, Kramers JD, et al. Geological setting and age of *Australopithecus sediba* from southern Africa. Science. 2010;328(5975):205–208. https://doi.org/10.1126/science.1184950

27. Barret JF, Keat N. Artifact in CT: Recognition and avoidance. Radiographics. 2004;24(6):1679–1691. https://doi.org/10.1148/rg.246045065

28. Courtney A, Coursey MD, Donald P, Frusch MD. CT and radiation: What the radiologist should know. Appl Radiol. 2008;37(3):22–29.

29. Van Grieken RE, Markowicz AA. Handbook of X-ray spectrometry: Methods and techniques. New York: Marcel Dekker; 1993.

Direct environmental impacts of solar power in two arid biomes: An initial investigation

AUTHORS:
Justine Rudman[1]
Paul Gauché[2]
Karen J. Esler[1]

AFFILIATIONS:
[1]Department of Conservation Ecology and Entomology, Stellenbosch University, Stellenbosch, South Africa
[2]Department of Industrial Engineering, Stellenbosch University, Stellenbosch, South Africa

CORRESPONDENCE TO:
Justine Rudman

EMAIL:
justine@sun.ac.za

DATES:
Received: 11 Apr. 2017
Revised: 14 July 2017
Accepted: 28 July 2017

KEYWORDS:
renewable energy; environmental impact assessment; Nama-Karoo; Savanna; mixed-method

HOW TO CITE:
Rudman J, Gauché P, Esler KJ. Direct environmental impacts of solar power in two arid biomes: An initial investigation. S Afr J Sci. 2017;113(11/12), Art. #2017-0113, 13 pages. http://dx.doi.org/10.17159/sajs.2017/20170113

ARTICLE INCLUDES:
✓ Supplementary material
× Data set

FUNDING:
Centre of Renewable and Sustainable Energy Studies (Stellenbosch University); National Research Foundation (South Africa)

According to recent national energy plans and policy documents, the number of renewable energy developments is expected to increase in South Africa, thus contributing to the diversification of the country's energy system. Consequently, numerous solar power developments are being deployed in the sunny arid interior – areas generally represented by the Nama-Karoo and Savanna Biomes. These developments come with a range of novel environmental impacts, providing opportunities for multidimensional exploratory research. Here, a mixed-method approach was used to identify and investigate possible environmental impacts associated with two types of solar power plants: concentrating solar power and photovoltaic. Structured interviews conducted with experts and experienced professionals, together with observations from site visits generated complementary findings. In addition to the risk of cumulative ecological impacts associated with individual solar plant developments, landscape impacts of multiple power plants and the direct impact on avifauna were found to be the most significant environmental impacts. These direct impacts appear to be most significant during the construction stage, which represents an intensive 10% of the total power plant lifespan. This investigation provides an early, broad and informative perspective on the experienced and expected impacts of solar power in South African arid regions as well as insights to possible future research areas.

Significance:

- Solar power represents a large component of the needed diversification of South Africa's electricity system.
- Research on the environmental impacts of solar power developments in the arid biomes of South Africa still is relatively scarce.
- Increased energy developments in the arid biomes will require knowledge of the associated impacts for conservation planning.
- Identification of environmental impacts throughout solar power lifespans enables informed management.

Introduction

The arid biomes of South Africa host an increasing number of renewable energy projects as the country diversifies its primarily coal-dependent energy system.[1] Generation capacities have been allocated to both concentrating solar power (CSP) and photovoltaic (PV) developments in the Integrated Resource Plan (IRP) of 2010[2] and the draft IRP update (IRP Update) of 2013, with solar power expected to represent an increasingly significant portion of renewable energy capacity[3]. Furthermore, the Renewable Energy Independent Power Producers Procurement Programme (REIPPPP) has awarded multiple CSP and PV projects, of which the majority are located in the arid Nama-Karoo and Savanna Biomes.[4]

Environmental impacts associated with individual solar power projects are currently being assessed through environmental impact assessments (EIAs) as governed by the *National Environmental Management Act* (*Act no 107 of 1998*).[5] Internationally, there is increasing research on the environmental impacts of solar power, but the scope of such enquiries varies widely. Previous investigations on the impacts of the solar energy technologies include: life-cycle assessments[6,7], studies on the landscape transformation and land-use efficiencies of development footprints[8,9], water use requirements[10,11], biodiversity and ecological implications[8,12,13], and a further specific focus on impacts to avifauna[14,15]. These impacts could vary per technology type (e.g. parabolic trough plants, central receiver plants) and the associated power plant design.[13]

A strategic environmental assessment (SEA) has been the only study to investigate the environmental impacts on a larger geographical scale in South Africa than has been done through EIAs which focus on individual developments. This SEA also identified potential areas – dubbed renewable energy development zones (REDZ) – associated with optimal social, economic and environmental impacts from wind and PV power developments. Outcomes of a SEA are not mandated but are intended to guide the siting of projects and expedite environmental authorisations for PV and wind power developments. This SEA, however, excluded the identification of such areas for CSP[16] – leaving a gap in the information for this solar power technology type.

Considering the relative novelty of solar power developments in the arid regions of South Africa and the introduction of a potentially wide variety of environmental impacts, the study of these impacts presents a dynamic space for multidisciplinary research.[17] Exploration of this topic has been off to a slow start locally, but research conducted in other parts of the world can guide early research in the arid regions of South Africa.[10,13]

Focusing within the arid regions of South Africa, we identify and investigate initial direct environmental impacts from utility-scale CSP and PV projects located across the Nama-Karoo and Savanna Biomes. A mixed-method approach was followed, which included conducting structured interviews with experienced and knowledgeable

individuals from diverse, yet relevant, backgrounds and affiliations; site visits to selected solar power plants; and spatial analysis using publically available data. The methods and results presented here were extracted from a more extensive study on the same topic, which includes a literature review that comprehensively covers the relevant policy, technological characteristics and environmental impacts.[18]

Methods

Study area

The availability of good solar resources and capacity on the national transmission grid are determining factors for where solar power developments are located, which explains why almost two thirds (62.5%) of existing PV power plants are located in the Northern Cape, with the rest distributed over five provinces. The proximity of the Orange River is an important consideration, particularly for the location of CSP projects in the study area, because of the need for water in the power cycles; this need contributes to the limited geographical distribution of CSP projects. Approved EIA applications for solar power developments in the Nama-Karoo and Savanna Biomes are shown in Figure 1.

Data collection and analysis

Structured interviews

Structured interviews were conducted from February to May 2016. Criterion[19] and snowball sampling[20] were used to identify interviewees from seven expert groups adhering to at least one of the following minimum criteria: (1) experience with or knowledge of the EIA process in South Africa and (2) experience with or knowledge of the environmental impacts of solar power developments. Ethical clearance was obtained before the start of the interview process from the Departmental Ethics Screening Committee of the Department of Conservation Ecology and Entomology and from the Research Ethics Committee of Stellenbosch University (proposal number SU-HSD-001751).

An interview form, written and presented in English, was designed to obtain qualitative and quantitative data. Prior to the interview, participants gave written consent for their participation and use of data in the study. A total of 20 interviews was conducted; some interviewees responded for both CSP and PV ($n=5$) and others for only CSP ($n=14$) or PV ($n=11$). Conducting interviews in person was the preferred method, but where circumstances prohibited, interviews were conducted telephonically or via Internet video conference. Responses to interview questions were directly recorded as text in an electronic copy of the interview form during the interview, which was then used as the transcript for analysis. Based on the nature of the responses obtained in the different sections of the interview form, the data were captured in Microsoft Excel or directly into the Computer Assisted Qualitative Data Analysis Software (CAQDAS) Atlas.ti 7®[21] in preparation for analyses.

Qualitative data were subjected to content and thematic analysis using Atlas.ti and two cycles of coding of the responses, which were obtained from different sections of the interview form. Coding aims to define qualitative data and involves identifying and highlighting parts of text, pictures or recordings that resemble a similar theoretical or descriptive concept, which is referred to as *code*.[22] During initial or open coding, responses to certain sections of the interview form were selected, after which a code was linked to each quotation. Second cycle coding involved the categorisation of codes based on predetermined subjects for discussion and patterns that emerged from open-ended sections of the interview forms. After the categorisation of codes into sub-themes or categories, content analysis was done, to prepare for thematic analysis.[23]

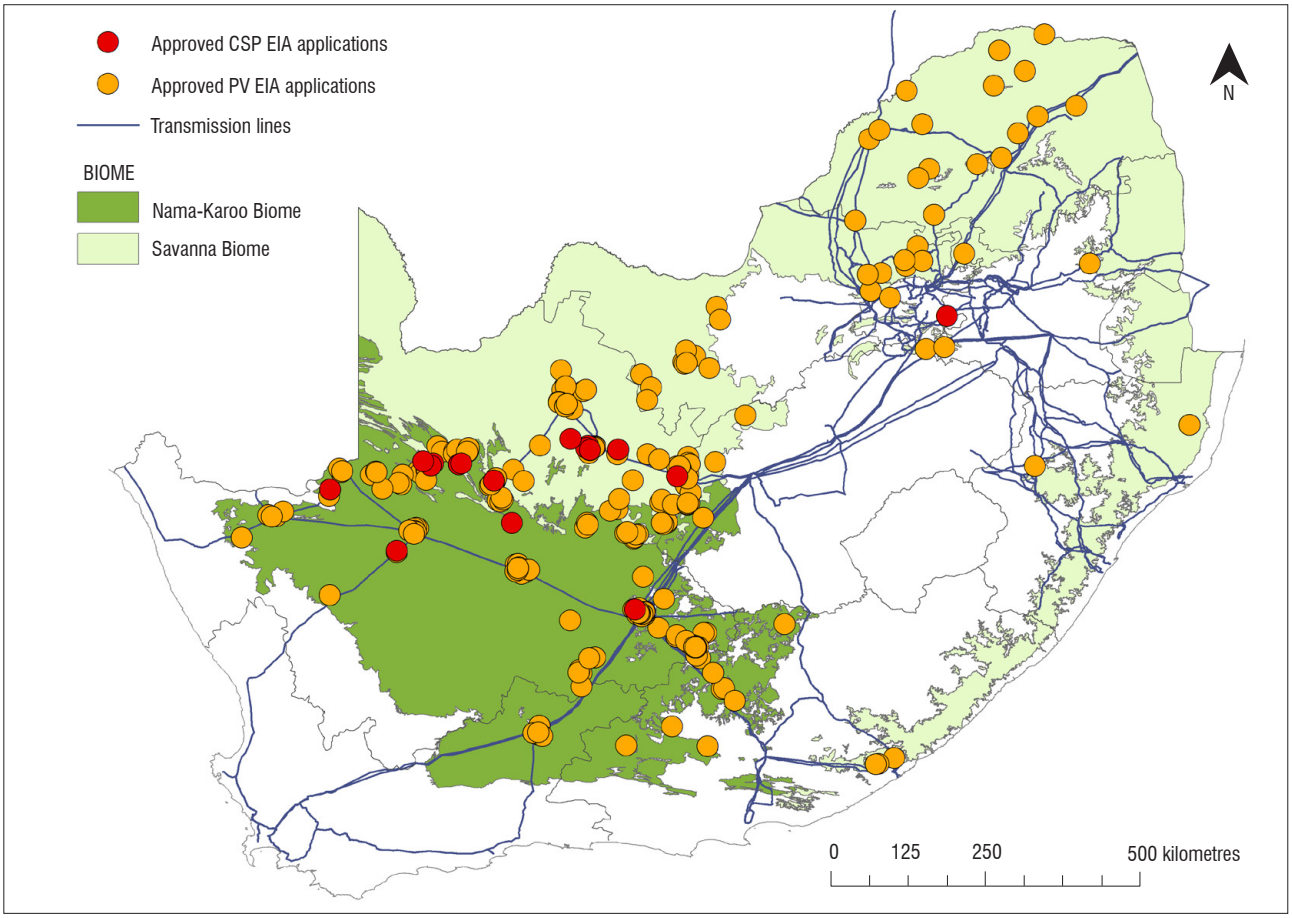


Figure 1: A map of the Nama-Karoo and Savanna Biomes. The red and orange areas show the distribution of approved environmental impact assessment (EIA) applications up to Round 3 of the Renewable Energy Independent Power Producers Procurement Programme for concentrating solar power (CSP) and photovoltaic (PV) developments, respectively. The national transmission grid is also shown.

Table 1: An explanation of ratings attributed to the severity and scale of impacts on different biophysical elements and solar power plant components

Rating	Severity of impact	Physical scale at which impact is incurred
0[a]	Interviewee unsure or regarded specific impact as irrelevant	Interviewee unsure or regarded specific impact as irrelevant
1	None	None
2	Light impact	Point specific (e.g. <1 km radius)
3	Moderate impact	Local ecosystem (e.g. 1–20 km radius)
4	Moderate to severe impact	Regional (e.g. 20–200 km radius)
5	Severe impact	National (across provincial boundaries)

[a]A score of zero was given by interviewees when they believed the impact was not relevant to the specific technology; these values were removed before calculating the median for each data subset.

Thematic analysis entails the discussion of categories and responses/codes within categories with the highest frequency of occurrence.

Quantitative analysis was limited to yes/no questions on the interview form and to a section in which ordinal data were obtained for ratings of the *severity* and *physical scale* of impacts during different stages of solar power developments. Ratings from zero to five were given, and their definitions are presented in Table 1. These ratings were given for the impacts of CSP and PV power plants on different biophysical elements and impacts from distinct solar power plant components on the biophysical environment as a collective.

Mann–Whitney U tests were used to compare ratings between the different stages of solar power development (construction and operation). The H_0 assumed no difference between different development stages for both comparisons of (1) the impacts on biophysical environment and (2) the impacts by different solar power plant components. The calculated *p*-values were compared at a probability level of 0.05 to test for statistical significance[24] in the ratings between the different development stages[25]. All statistical analyses were done using the Microsoft Excel statistical plugin, XLSTAT®.[26]

Site visits

Site visits were included to support the interview results within the context of the chosen biomes. The visited sites were widely distributed within these biomes, instead of focusing on impacts from individual power plants with limited distribution. Four PV plants and two CSP plants were visited in June 2016. The purpose of the site visits was to observe the status and environmental impacts of existing solar power developments through personal, on-site observations; interactive discussions; and photographic recording, where permitted. Hosts were informed of the purpose of the visit and agreed to share information accordingly. The agenda of a site visit included induction followed by a thorough tour of the development while discussions took place; each site visit lasted approximately 2–3 h per site. All notes that were made during the site visits were based on what was observed on site as well as on the experiences shared by site visit hosts. The locations of the power plants visited (Supplementary Figure 1) and descriptive information are included in the thesis on which this paper is based.[18] Experience data obtained from six site visits to solar power plants were thus interpreted in the context of each unique power plant, and no additional analysis was done.

Spatial data collection

A spatial data set that summarises EIA applications for CSP and PV (hereafter, the EIA data set) was obtained from the South African Department of Environmental Affairs (DEA). The EIA data set was used as the primary reference for identifying areas where CSP and PV developments are taking place as well as which biomes are being impacted. To obtain a regional understanding of the impacts experienced from these developments, the EIA data set was used to investigate impacts on the landscape by using a combination of topic-specific spatial data sets. The majority of these data sets are publically available and/or obtained from researchers in government departments with permission (a summary of the spatial data sets is given in Supplementary Table 1).

All spatial and geographical data analyses were conducted using the geographical information systems software package ArcGIS®. Appropriate tools from ArcGIS were used to manipulate and combine data sets as well as to extract information that reveals insight into the impact of solar power developments across the Nama-Karoo and Savanna Biomes, using the solar power EIA application areas as the starting point.

Results

Interview results

Interviewees were categorised into seven expert groups (Table 2), which served as a representative sample of the greater knowledgeable, qualified and experienced population of experts.[27,28] Not all interviewees disclosed their affiliations, but those who did were associated with one or more of the following entities: the Solar Thermal Energy Research Group at Stellenbosch University, BirdLife South Africa, Council for Scientific and Industrial Research, Eskom, the DEA, the South African National Energy Development Institute, World Wide Fund for Nature South Africa, Umvoto Africa (Pty) Ltd, the Plant Conservation Unit at the University of Cape Town, Simon Todd Consulting, Khi Solar One (Pty) Ltd (Abengoa), Golder Associates Africa (Pty) Ltd and Savannah Environmental (Pty) Ltd. The number of responses obtained from each expert group for CSP and PV is summarised in Table 2.

Table 2: A summary of the representation of the interviewees and the number of responses for the two different solar power technologies

Expert group	Concentrating solar power	Photovoltaic
Research entity	2	1
State utility	1	1
Designated authority	1	1
Registered environmental assessment practitioners	2	5
Representatives from independent power producers	1	1
Legislation/policy developers	1	1
Specialists	4	3

Through the coding and analyses of the interview data from all interviewees ($n=20$), responses were summarised into three prevailing themes (results for a fourth theme are excluded here).

Theme 1: Direct environmental impacts from solar power development

Interviewees were asked if they were aware of any adverse direct environmental impacts from solar power developments on the natural environment. To this question, 95% of interviewees ($n=19$) responded yes and 5% ($n=1$) responded no. Interviewees were then provided an opportunity to mention any known impacts related to solar power development. In total, 47 different impacts were coded in this section and grouped into seven biophysical impact categories, which are listed and described below:

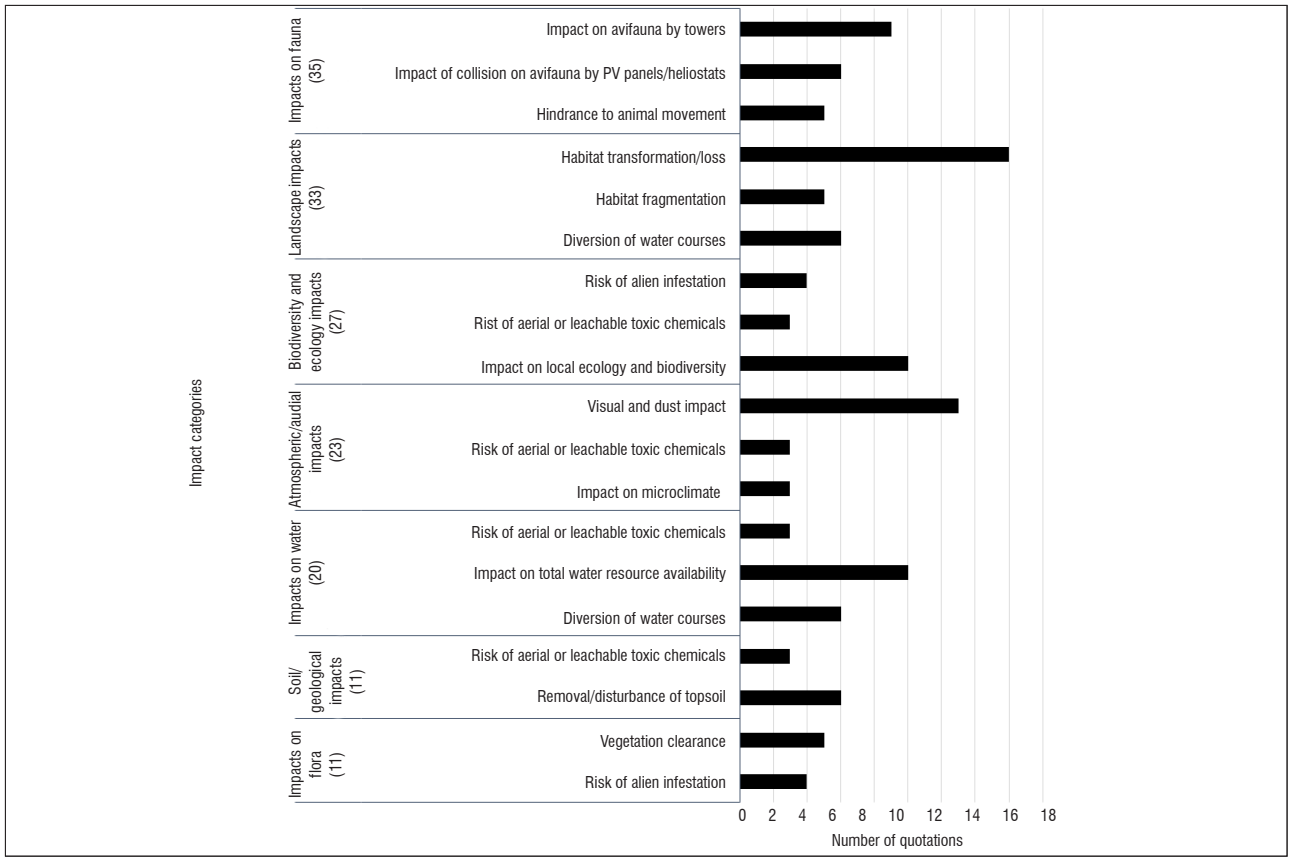
1. **Atmospheric and audial:** Impacts include changes in albedo, microclimate, audial impact, light pollution and visual impact.
2. **Biodiversity and ecology**: Impacts mentioned as *biodiversity* or *ecological* impacts and impacts with potential to have an effect on the dynamics between biological and physical ecological proxies.
3. **Fauna:** All mentioned impacts with specific relevance to animals.
4. **Flora:** All mentioned impacts with specific relevance to plants or vegetation.
5. **Landscape:** Impacts on the land which transcend the boundaries of a development or refer to the impact of a development on the landscape.
6. **Soil and/or geological impacts:** Impacts by solar power developments on soil and/or the underlying geology.
7. **Water:** Resource quality and size-related impacts for both surface- and groundwater resources.

The frequencies with which the most common impacts were mentioned within the seven biophysical impact categories are shown in Figure 2. The impact on fauna with a particular focus on avifauna was found to be the most prominent impact category associated with solar power developments, followed by landscape impact and impacts on biodiversity and ecology. Of a seemingly lower concern, impacts on flora and soil or geological impacts were mentioned less frequently during the interviews. Within the given impact categories, specific impacts such as habitat transformation or loss, visual and dust impacts and impact on total water resource availability were frequently recorded.

Although the impact on water resources was not as frequently coded as that of impacts on fauna, the long-term impact of multiple CSP projects on water availability was a recorded concern for at least three interviewees. Regarding this concern, an interviewee who manages a team of environmental assessment practitioners (EAPs) shared that the water in the Orange River is largely allocated to other uses, and current predictions are that the Orange River is less than 10 years away from not being able to meet further development needs.

Although general impacts of solar power were recorded, interviewees were at liberty to mention specific impacts related to CSP or PV. Of these, the impact on avifauna from central receiver towers and avifauna collision impacts with PV panels or heliostats from central receiver plants were found to be the most frequently mentioned. The risk of toxicity of thermal oil used in parabolic trough plants and PV panels was also mentioned.

Regarding the numerical ratings obtained from interviewees, several interviewees commented that the ratings they gave (based on the scores and definitions given in Table 2) included the assumption that the needed management actions or plans are in place, i.e. if management plans were not in place, a higher rating might have been given.



PV, photovoltaic

Figure 2: Summary of the seven biophysical impact categories in descending order by total number of quotations per category as indicated in brackets. Biophysical impacts that were mentioned more than twice per impact category are listed per impact category. A quotation represents a single event in which the specific impact was mentioned.

The only CSP biophysical environment impacts (shown in Supplementary Figure 2) rated as having higher severity during operation were those on surface water usage and quality; birdlife; and visual impairment. The median rating for the severity of impacts on all other biophysical elements during construction was equal to or higher than that of operation. Ratings for the physical scale of impacts on these biophysical elements were found to be similar during the two stages in almost all cases except for groundwater quality, for which the median rating was higher during construction. Interestingly, for visual impact, the range of ratings for the physical scale of the impact was found to be the highest (minimum = 2, maximum = 4) during operation, indicating the uncertainty of the actual visual impact during CSP plant operation. The mean ratings for PV developments (also shown in Supplementary Figure 2) also indicate that the severity of impacts is the same for both stages or is higher during construction, again with the exception of visual impact which received a higher rating for the operational stage. For both CSP and PV, the highest severity ratings were received for the impacts on soil, vegetation and increased dust during construction.

The medians and ranges (minimum to maximum) of the ratings obtained were calculated for the severity and physical scale of the impacts from the various power plant components of CSP and PV power plant development (shown in Supplementary Figure 3). For CSP, the median rating was 2 for the majority of the power plant components for both severity and physical scale, with the exception of roads and the solar field for which impact severity was rated higher during both stages. The range of ratings received for the physical scale of impacts by substations and/or power lines was the largest of all power plant components for both development stages. The findings for the impacts by the various components of PV developments are similar to those of CSP developments, with the exception of evaporation ponds and energy storage facilities which are irrelevant for PV. The severity and scale of impacts by roads, substations and/or power lines and the solar field of PV developments were found to be generally higher than those of other power plant components during the construction stage. Components such as energy storage facilities, offices or on-site accommodation and temporary structures or scaffolding had the narrowest rating ranges (minimum = 1, maximum = 2; or minimum = 2, maximum = 3), indicating that the impacts from these components are fairly contained to the development footprint.

Table 3 summarises the *p*-values for (1) biophysical elements and (2) power plant components, for which a significant difference was found between the construction and operational stages. Here we see differences in impact severity between these two stages on almost all biophysical components for PV developments. No significant difference was found for the rated physical scale of biophysical impacts between the two development stages for PV. However, strong evidence of a significant difference between development stages was found for at least the physical scale of dust- and vegetation-related impacts from CSP developments. The results for the ratings of the power plant components show little consistency between CSP and PV developments. The only commonality is a significant difference between the two development stages in the impact severity of temporary structures.

Theme 2: Feedback and experience with EIA process

Feedback and comments regarding the current EIA process and the coverage of impacts from solar power projects in EIAs are given in this theme. Many responses to this section included suggestions for amendments to the EIA process and/or suggestions for minimising and managing impacts; the latter are not included here.

Further to listing impacts related to solar power developments (as indicated in Theme 1), interviewees were asked whether they think EIAs sufficiently cover all impacts of a project on the biophysical environment; the majority of interviewees replied yes (n=11), one interviewee was too unsure to answer, and the rest replied no (n=8). Three interviewees explicitly stated that all impacts are covered in detail. Some interviewees furthered their response with a comment, and those comments mentioned more than once are summarised in Figure 3.

Although the majority of the interviewees agreed that the current EIA process sufficiently covers all impacts of solar power developments, two of the most common responses, indicated in Figure 3, highlight EIA implementation as a key concern. These comments suggested that the aspects generally omitted in the EIA process are 'cumulative impacts' and 'analysing topsoil and vegetation removal in depth'. Two specific examples of negligible assessment of cumulative biophysical impacts were atmospheric pollution and insufficiently investigated resource requirements (e.g. water) prior to the start of a development.

Table 3: The biophysical components and power plant components for which there was a significant difference in ratings between construction and operation. These are given for severity and physical scale for both concentrating solar power and photovoltaic. Significance (*p*-)values are provided in parentheses[a].

Rated subject	Concentrating solar power[b]		Photovoltaic[c]	
	Impacts on biophysical environment[b]	**Impacts by different power plant components**	**Impacts on biophysical environment[b]**	**Impacts by different power plant components**
Severity	Soil (0.033) Mammals (<0.001) Reptiles (<0.001) Vegetation (0.004) Dust (0.003)	Waterworks (0.009) Temporary structures/scaffolding (0.019)	Soil (0.002) Air quality (0.033) Birdlife (0.010) Mammals (0.001) Reptiles (<0.001) Vegetation (0.004) Audial impact (<0.001) Dust (<0.001)	Roads (0.039) Solar field (0.002) Offices/on-site accommodation (<0.001) Temporary structures/scaffolding (0.001)
Physical scale	Vegetation (0.011) Dust (0.009)	Power block/inverter block (0.028) Energy storage facilities (<0.001)		Temporary structures/scaffolding (0.039)

[a]*Results from Mann–Whitney U test,* n=*15,* p-*level (alpha level) used = 0.05.*

[b]*In all these cases, construction ratings were higher than those for operation.*

[c]*With the exception of the power block/inverter block, all the ratings were higher for construction than for operation.*

An interviewee with prior practical experience of EIAs at solar power developments specifically mentioned the invasion of alien flora and the attraction of fauna to evaporation ponds as impacts that are not covered sufficiently during the operation stage of a power plant. Five interviewees said they think the impacts are described in detail for both development stages and/or no impacts are omitted in the EIAs.

Although EIA implementation was shown to be a concern, further comments indicate that, when properly implemented by competent EAPs and environmental consultancies, the EIA process is sufficient. Two respondents specifically commented that the legislation is sufficient, but implementation thereof and the follow-through from EIA to the environmental management plan from a legislative perspective during construction might be a weak area. An EAP from the Council for Scientific and Industrial Research commented that 'EAPs have a good understanding of impacts, but the assessment thereof is not reinforced by site visits'. An interviewee who has experience as an EAP and as a specialist commented positively on the thoroughness of the DEA to intervene when there is suspicion that an EIA may have been insufficiently completed.

An employer of an independent power developer with previous experience as an EAP described the central receiver plant, Khi Solar One, as a 'first child' from which many valuable lessons have been learnt. This response is similar to that from an employer at the DEA who openly stated that some of the impacts that might have been missed in the earlier projects' EIAs are a matter of 'learning as we go'.

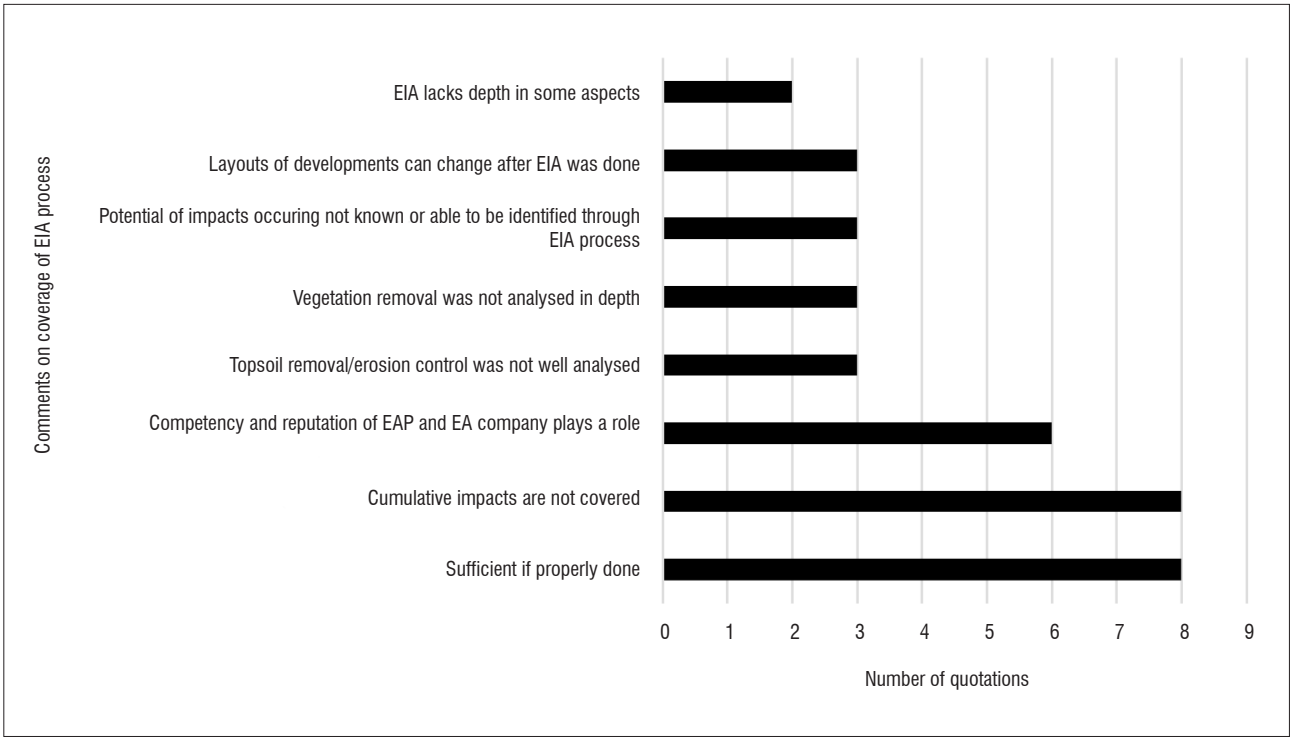
During early stages of project planning and the EIA scoping phase, preliminary impacts of solar power plants are identified based on spatial biodiversity data sets. The quality and representativeness of these data sets are relevant to minimising impacts on the underlying biodiversity at a specific location. Questions about the biodiversity data sets used for this purpose were included in the interview form. Almost all interviewees (85%; $n=17$) claimed to know which data sets are being used for baseline studies prior to solar power developments. Although not all interviewees could recall the correct data set names, a trend was noticed in the frequently mentioned data set topics. The South African National Biodiversity Institute, the South African Department of Water Affairs and Sanitation, and the DEA were the three data sources mentioned most frequently. Interviewees who knew which data sets are being used were asked if existing field survey archives, spatial data sets and maps were sufficient to predict the impact of solar power developments in South Africa. To this question, 41% ($n=7$) replied yes and 59% ($n=10$) replied no. These responses were furthered by comments, amongst which the following three points were most frequently recorded:

- Current data sets and maps have insufficient resolution and/or are outdated, especially in arid regions, and need to be updated ($n=14$).
- Ground-truthing is necessary (verification of features represented in a spatial data set with field investigation) ($n=10$).
- A more strategic, tiered, systematic and cooperative approach is needed to keep data sets updated ($n=4$).

An EAP from the Council for Scientific and Industrial Research confirmed that '[m]uch of the information used in solar power EIAs has become generic, and should be more pertinent and relevant to the development site.' In a related comment, an interviewee with experience as a specialist in EIA application suggested that a mandatory requirement to submit field data collected for EIA purposes to the South African National Biodiversity Institute after a certain time period could aid in keeping national data sets updated.

Theme 3: Reference to SEA process

Throughout the interview process, mention was made of the SEA that was completed for wind and PV power. The feedback about the SEA process and the linkage to EIAs was limited to three specific points: (1) a perception that the outcomes of the first wind and solar SEA are not utilised to guide EIAs; (2) a view that the usefulness of the SEA is limited given that the distribution of renewable energy projects is in reality constrained by the existing transmission grid infrastructure; and (3) a suggestion that the SEA process must be improved and that CSP should be included in the new SEA being performed for PV and wind power.



EAP, environmental assessment practitioner; EA, environmental assessment

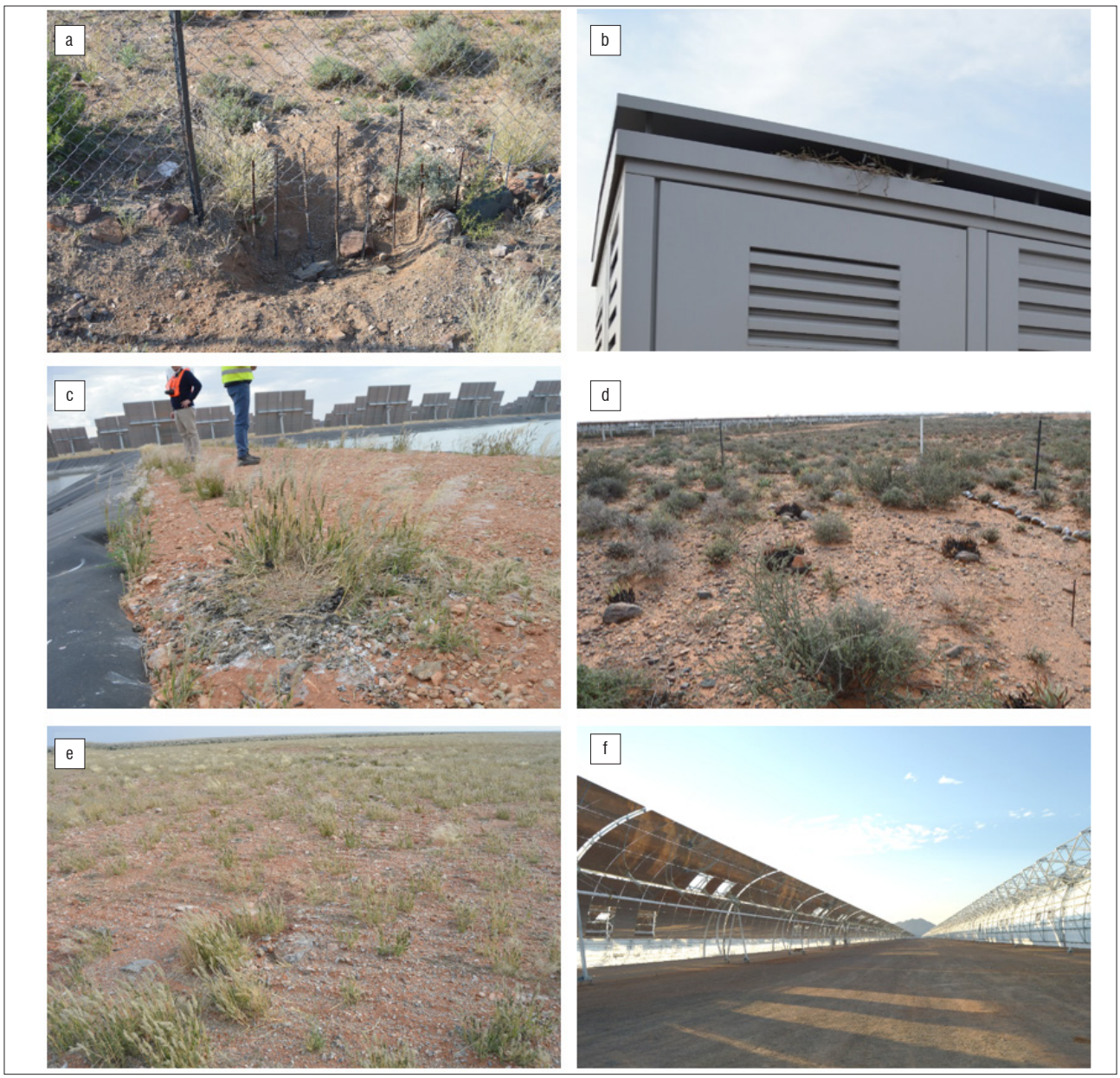
Figure 3: Summary of similar comments in response to the question of the sufficiency of the environmental impact assessment (EIA) process to capture all possible environmental impacts of a project.

Site visits

The observations, conversations and photographs recorded during the field trip were combined to offer a collection of in-field experience to support the results from the interview process. These findings largely correlate with the interview results, but site visits allowed for more specific insights regarding matters such as animal interactions and water impacts. With the exception of waste materials at sites still under construction and hydraulic fluid spills at Khi Solar One (Site 5), no unexpected adverse environmental impacts were observed during the site visits. Key findings from the site visits are summarised in Table 4 and supported by photographs of specific phenomena, presented in Figure 4, observed at different sites.

Spatial analysis

Areas under solar power development were extracted from the renewable energy EIA applications database for all projects up to the third round of the REIPPPP. Power plants with approved EIA applications were differentiated from those selected as preferred bidders of the REIPPPP. The projects with approved EIA applications comprise a larger surface area than those of the preferred bidders, highlighting how many EIAs have received approval in the study area. The preferred bidders are a limited number of projects committed for construction and operation and thus contribute to the cumulative direct environmental impacts. A subset of spatial data for solar power developments (all of which are approved solar power EIAs and preferred bidders) was used to quantify the cumulative affected area in the different biomes, vegetation types and other topic specific land uses.



Photos a–f: Justine Rudman

Figure 4: Photographs of key observations made during site visits (as reported in Table 4). (a) An example of how animals burrow underneath the development fence and an improvised attempt to keep them out at Site 3. (b) A nest in a small opening at the top of a transformer building at Site 4. (c) An empty nest at the edge of an evaporation pond at Site 5; some of the heliostats of the solar field are visible in the background. (d) One of the 'green areas' at Site 4 where six kraal aloes were relocated from the solar field prior to construction. (e) The topsoil embankment at Site 5 during the early stages of rehabilitation. (f) A row of parabolic troughs at Site 6 showing the cleared and compacted ground of the solar field.

Table 4: Selected key findings from the site visits per impact category. Findings are arranged as associated with either construction or operational activities.

Impact category	Observations and findings related to construction activities	Observations and findings related to operational activities
Impacts on fauna	All sites were different in the way in which animal movement into and out of the development footprint was allowed or managed (see Figure 4a). Steenbok (*Raphicerus campestris*), rodents and snakes such as puff adders (*Bitis arietans*) were said to be common occurrences within development footprints.	
	Birds nesting in power plant infrastructure was recorded at Sites 1–4 (see Figure 4b). A striped polecat (*Ictonyx striatus*) with rabies had been found near the temporary buildings of Site 4.	Rodents and aardvark (*Orycteropus afer)* were said to have gnawed on cables and wires at Sites 1 and 2. Birds had been observed in flight (e.g. falcons, eagles, flamingoes) and nesting (black-winged stilts) around and in the evaporation ponds of Sites 5 and 6 (see Figure 4c). Three flux-related bird fatalities had been recorded at Site 5 (central receiver facility). Two mammal drownings had occurred at Site 6: bat-eared fox (*Otocyon megalotis*) and an aardwolf (*Proteles cristata*).
Impacts on flora	Vegetation was removed in the solar fields of Sites 1, 5 and 6 but kept intact at the other sites. At Site 4, two green areas were established for the relocation of six kraal aloes (*Aloe claviflora*); see Figure 4d.	Where natural vegetation was kept intact, it was seen as an effective natural dust suppressor. At Site 2, the vegetation was kept intact and the development footprint was also still used by the landowner for grazing sheep. Vegetation regrowth was generally encouraged at all visited sites, except for Site 6 where vegetation in the solar field was considered a fire hazard. Alien species such as Mexican poppies (*Argemone mexicana*) and prosopis (*Prosopis juliflora)* were recorded at Sites 3 and 5, respectively.
Soil/ geological impacts	Topsoil clearance had occurred in the solar fields of Sites 1, 5 and 6 but was rehabilitated at an embankment (see Figure 4e). Soil was impacted at all sites by the construction or installation of pylons, trenches and roads.	The entire solar field at Site 6 needed to be levelled on different terraces and the soil compacted (see Figure 4f). Depending on the stormwater management plan, erosion was a problem at some of the sites.
Impacts on water	The stormwater management plans were problematic at all sites and required revision.	
	Water was predominantly used for dust suppression during construction.	No standard practice was found regarding the regularity of photovoltaic (PV) panel washing. This frequency ranged from once every 6 weeks to twice a year. Borehole water was used at Sites 1, 2 and 3, and the treatment varied between sites. No quantities were given for this activity. Sites 5 and 6 had annual water use permits of 300 000 m^3 and 400 000 m^3, respectively.
Aerial/ audial impacts	Dust and noise were the only recorded aerial or audial impacts. During construction of Site 5, complaints were apparently received about the excessive dust at a small community 5 km away.	Dust during strong winds was found to be problematic at sites where vegetation and topsoil were removed. Eventual regrowth in the solar field assisted in this regard.
Spills and waste	Concrete spills were noticed at Site 4. Excessive waste (e.g. plastic, pallets and broken panels) and spills were recorded in the construction camp of Site 3 even though this area was supposed to have been rehabilitated.	Oil or hydraulic fluid spills were recorded to have occurred at four of the six visited sites. The containment and treatment of these varied per site. Lessons were learnt at Site 6 regarding salt spillage and leakage of the heat transfer fluid at the neighbouring Kaxu Parabolic trough plant.

The nearest town and associated technology of the visited sites were as follows: Site 1, Touwsriver, concentrated PV; Site 2, Hanover, PV; Site 3, De Aar, PV; Site 4, Copperton, PV; Site 5, Upington, CSP (central receiver); Site 6, Pofadder, CSP (parabolic trough). Full site visit notes and findings from the study scope are included in the thesis on which this paper is based.[18]

Table 5: The area per biome for which solar power environmental impact assessment (EIA) applications have been approved and the total area of projects that were preferred bidders throughout the first three rounds of the Renewable Energy Independent Power Producers Procurement Programme (REIPPPP)

Biome	Total area[a] of approved EIA applications per biome (km^2)	Percentage of the total area[a] with approved EIAs per biome	Total area with approved EIA application for preferred bidders[b] (km^2)	Percentage of total area for preferred bidders[b]
Nama-Karoo Biome	4455.0	49.3%	702.0	70.3%
Savanna Biome	2854.1	31.6%	228.2	22.9%
Grassland Biome	988.9	10.9%	16.7	1.7%
Fynbos Biome	257.0	2.8%	30.4	3.0%
Succulent Karoo Biome	234.4	2.6%	5.1	0.5%
Azonal Vegetation	176.4	1.9%	10.6	1.1%
Albany Thicket Biome	68.4	0.8%	–	–
Desert Biome	5.4	0.1%	5.4	0.5%
Indian Ocean Coastal Belt	0.9	0.01%	–	–
Total	**9040.4**	**100%**	**998.4**	**100%**

[a]Total area for all approved solar power developments throughout Rounds 1–3 of the REIPPPP. Not all of these projects continued on toward construction and operation.

[b]A subset and smaller area than that of all approved EIAs.

Impacts on biomes and vegetation types

At the time of writing, projects had been approved for Round 3.5 and Round 4 of the REIPPPP, but the EIA data of these projects were not included in the latest data sets made available by the DEA at the time of analysis. Table 5 summarises the total area per biome for which (1) solar power EIA applications have been approved and (2) projects have been assigned to preferred bidders throughout the first three rounds of the REIPPPP. These results confirmed that the Nama-Karoo and Savanna Biomes have a clear majority proportion of area (70.32% and 22.85%, respectively) under development by preferred bidders' projects. In addition to the proportional impacts on biomes and vegetation types, land-cover data revealed that an estimated 95% of the area under solar power development was previously classified as 'low shrubland' and 'bare ground'.[29]

The total area of land under solar power development per vegetation type[30] within the Nama-Karoo and Savanna Biomes was calculated. (A summary of the ten most affected vegetation types is presented in Supplementary Table 2). Indicating the total area per vegetation type in parentheses, the five most affected vegetation types by development footprints of preferred bidders were Bushmanland Arid Grassland (256.3 km^2), Northern Upper Karoo (153.9 km^2), Kalahari Karroid Shrubland (128.3 km^2), Bushmanland Basin Shrubland (116.3 km^2) and the Gordonia Duneveld (93.5 km^2). The Bushmanland Arid Grassland and the Eastern Upper Karoo are the most targeted vegetation types for solar power development, and represent 26% and 15% of the preferred bidders' area, respectively. However, both types have a conservation status of 'least threatened'.[30]

Impact on protected areas and biodiversity planning areas

The likelihood that new developments are located in already-existing protected areas was assumed to be less than that in areas not yet protected, although areas of ecological significance have already been earmarked through the National Protected Area Expansion Strategy (NPAES) last updated in 2010.[31] NPAES areas were allocated and identified by South African National Parks as intact and unfragmented areas of high importance for ecological persistence and biodiversity representation and suitable for the creation or expansion of large protected areas.[31] It was found that the only areas overlapping with NPAES focus areas were those with approved EIA applications for PV developments and no preferred bidders. NPAES areas are shown in relation to the EIA applications areas in Figure 5.

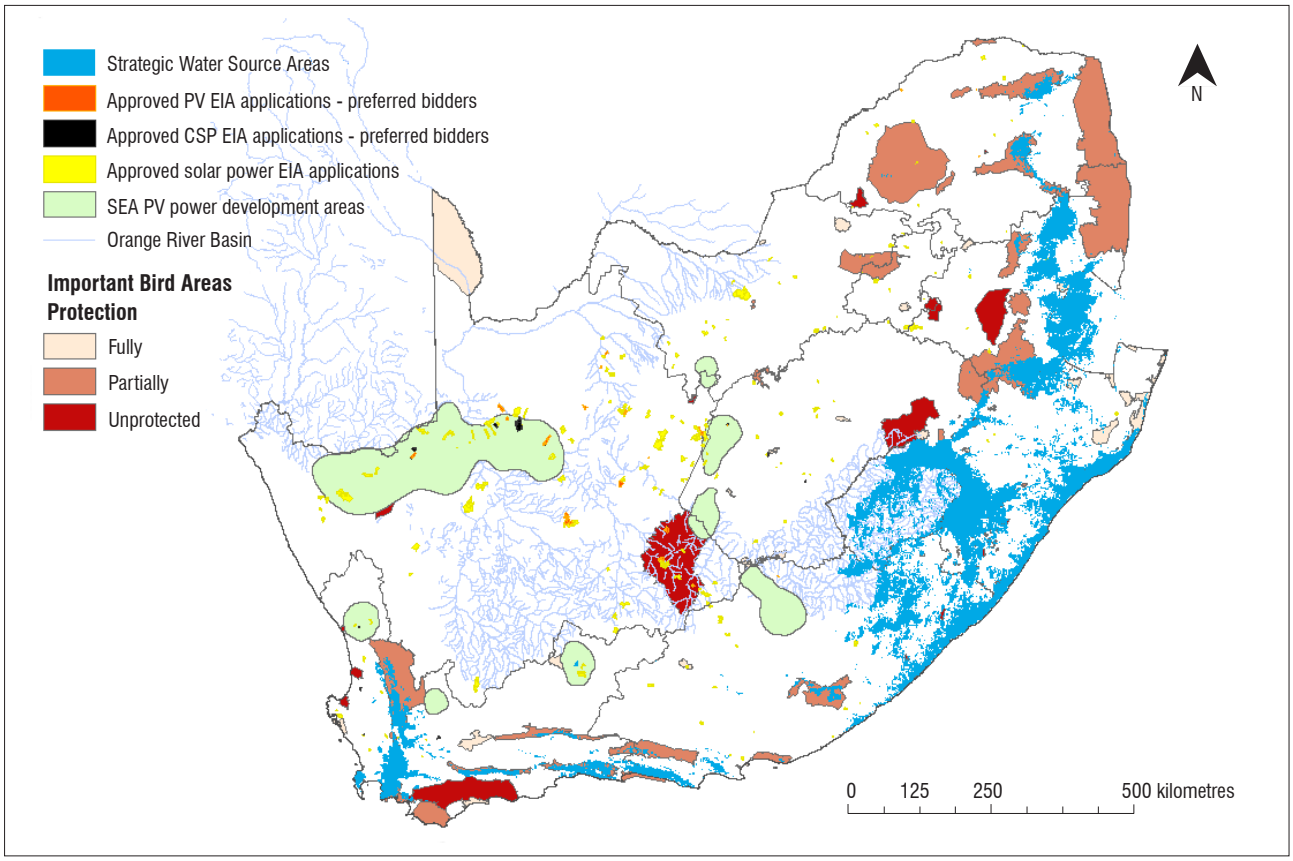
Important Bird Areas (IBAs), as identified by BirdLife South Africa[32], are classified as 'unprotected', 'partially protected' or 'fully protected' (Figure 5). A subset of the IBAs, which are located within the Nama-Karoo and Savanna Biomes, was analysed to determine what proportion of these areas has approved solar EIA applications and preferred bidders' developments located within them. No preferred bidders are located within fully protected IBAs, but 7.6 km^2 and 168.8 km^2 of CSP and PV developments, respectively, are located in unprotected IBAs. The unprotected IBAs that were identified as being affected by PV preferred bidders' developments in the Northern Cape are the Platberg-Karoo Conservancy, Mattheus-Gat Conservation Area and, in the North West Province, the partially protected Magaliesberg IBA. The unprotected IBA, Mattheus-Gat Conservation area, was identified as being affected by CSP developments of preferred bidders. Proximity of EIA applications to IBAs was also calculated, and it was found that approximately 88% of solar power developments are more than 10 km away from any IBA.

Overlap of CSP development areas and strategic water source areas was investigated, but no overlap in location was found (Figure 5).

Lastly, the areas identified with top PV development potential throughout the National SEA for wind and PV power as REDZ[33] were analysed to determine how many approved EIA applications and preferred bidders' developments are located within these areas. Of the approved EIA applications for PV, 17% were located within the PV REDZ, and 8% of these projects were selected as preferred bidders. Of the total preferred bidders for PV developments, only 15% were located in a PV REDZ. The co-location of these areas with those of approved solar power EIA applications can be seen in Figure 5.

Footprint and distribution at increased capacity allocations

Capacity allocated to solar power projects in the REIPPPP represent approximately 19% and 18% of the allocations to PV (1899 MW of 9770 MW) and CSP (600 MW of 3300 MW) in the IRP Update Base Case scenario, respectively.[3,34] Potential future affected areas can be calculated on the assumption that the land-use efficiency of these technologies remains constant and that the remaining 81% for PV and 82% for CSP will be located in similar and adjacent solar resource areas.



CSP, concentrating solar power; SEA, strategic environmental assessment

Figure 5: A map of South Africa showing national strategic water source areas and photovoltaic (PV) power Renewable Energy Development Zones. Areas with approved solar power environmental impact assessment (EIA) applications and the following protected or sensitive biodiversity areas are shown for the Nama-Karoo and Savanna Biomes only: Important Bird Areas, South Africa's Protected Areas and National Protected Areas Expansion Strategy focus areas.

The placement of the expected developments can be extrapolated to similar biophysical areas; that is, the variation of proportional distribution of projects within biomes may be minimised as indicated in Table 5. Limitations in these assumptions include unknown timing of when the area would be transformed, the unknown extent of transmission grid expansion and the assumption of consistent land-use efficiency for both CSP and PV projects. This extrapolation and calculation revealed that potential future areas under solar power development per biome will likely be relatively low, with the expected transformed footprint by 2030 at approximately 1.57% in the Nama-Karoo and 0.31% in the Savanna Biome.

Discussion

Studies of public perception and attitudes towards renewable energy technologies are present in the literature[35-37], but experience of these technologies from professionals in the field is not as easy to find. Spatial analysis has been used for questions related to land-use efficiency of solar power developments[38,39], but little work exists apart from the work of Fluri[40] and the guidance provided through the identification of REDZ in the SEA completed for wind and PV power[33]. This study is an attempt at identifying the direct impacts of South African solar power developments.

Synthesis of findings

Most recorded responses from the interviews regarding the various impact categories, such as impact on avifauna and water consumption, are similar to the findings presented in a recent review paper by Hernandez et al.[13] and a more detailed earlier overview presented by Tsoutsos et al.[10] Numerical ratings further supported the findings from the content analysis (e.g. high median rating for the impact severity on birdlife and water usage during operation by CSP developments). Impacts similar to those assessed by Turney and Fthenakis[12] – such as impacts to wildlife, wildlife habitat and geohydrological resources – were also recorded. The interview results contributed valuable insights during this first study within the Nama-Karoo and Savanna Biomes. However, the method and scope of this study provides a starting point to a multitude of potential future studies with deeper and narrower scopes in the same study area. Because of the widely acknowledged reality of these impacts, the study arguably provides an opportunity for context-specific description and management guidelines of these impacts within an ecological context of a development.

The concern for the impact of CSP on water resource availability together with the comment that cumulative impacts are not covered sufficiently in the EIA process highlights the need for strategic planning of water resource allocation to CSP. This need is specifically relevant around the Orange River Basin where water supply is known to be limited[41]; strategic planning of this kind is in accordance with the term at which power purchase agreements are signed with developers. Early acknowledgement of a possible risk of over-allocation of water resources provides a starting point for proper description and management of these impacts resulting from CSP developments around the vicinity of the Orange River.

In addition to the added value derived from observing impacts 'on-the-ground' during the site visits, the positive attitudes and practice at some sites were unexpected, pleasant findings. As mirrored in Theme 1 of the interview results, dust impact associated with vegetation and topsoil removal in the solar field was regarded as a major impact within the immediate environment of a solar power plant. Impacts associated with

fauna, flora and water mostly appeared to be well planned for during the EIA phase, and when unanticipated impacts occurred – such as the nesting of birds in structures or buildings and the attraction of species to evaporation ponds – impromptu actions were implemented. These observations correlate with feedback on the coverage of the EIA process in which it was highlighted that the attraction of certain species to development footprints is not sufficiently covered in the EIA process. Although monitoring data collection is needed to confirm, this scenario is similar to the 'mega-trap' concept described by Kagan et al.[15], in which solar power developments act as an ecological trap as a result of the creation of favourable areas which offer reproductive and foraging advances within the surrounding ecosystem.[42]

Exploring the impacts on fauna and habitat transformation in general combined with the feedback regarding the quality of biodiversity data sets used in the early stages of the EIA process justifies a clear concern, but it also creates a timeous opportunity. The timing in South Africa is ideal, considering that renewable energy developments are still in the early stages and there is a high potential to gain experience in these impacts. The opportunity lies in focusing on mapping the biodiversity, including specialist studies in planned development areas and developing best-practice guidelines that can proactively avoid impacts on species diversity as well as take into account seasonal migration of avifauna. The suggestion from interviewees that there be a strategic, cooperative approach to keeping these data updated is thus a relevant one. Furthermore, there may be significant potential to update data sets using the in-field data, which gets collected as part of the EIA process.

In addition to ensuring the use of representative data sets to avoid adverse impacts, identified direct impacts would be best supported with species-specific monitoring data to determine specific impact(s) on avifauna from solar power development. No such data were published or available to the public at the time of writing. Such monitoring data would support the impacts previously mentioned that are specific to CSP and PV, of which 'impacts on avifauna by CSP towers' and 'collison impact by PV panels or heliostats' were the most popular; these impacts are reflected in international studies as well.[8,14,15] Furthermore, monitoring data of this nature could inform management and mitigation measures, regulations and the establishment of IBAs as has been done by BirdLife South Africa.[32,43] However, no evidence based on South African data was available for any of these impacts as peer-reviewed studies.

Landscape outlook

The various environmental impacts associated with a single solar power development could be insignificant, but the landscape-wide accumulation of impacts is a possible concern. The findings from the interviews and site visits enabled us to identify a spectrum of impacts that occur at the solar power plant level; the spatial analysis assisted in investigating the distribution of these impacts across the Nama-Karoo and Savanna Biomes. All three of these data collection techniques and the findings of impacts on fauna, biodiversity and ecology, and landscape transformation link back to the importance of appropriate siting and mapping. The extreme importance of representative biodiversity data should thus be heavily stressed.[44,45] Missing data risks putting in danger individuals and/or populations of species in ecosystems with limited geographical distributions as well as affecting the alpha and even beta diversity.[46]

An aspect weakly investigated in this paper is the assessment of the impacts from support infrastructure (e.g. power lines and access roads), which could result in widespread habitat transformation. Including analysis of such infrastructure spatial data, in combination with real incident data, would give further insight to the expected ecosystem-level landscape-scale impacts of supporting transport and access infrastructure associated with solar power plants.[47,48]

The motivation behind the wind and PV SEA is good in that it aims to identify areas as REDZ where significant adverse environmental impacts are minimised. Given that the SEA is based on national and local biodiversity data sets[33], maintaining updated data sets is critical to guide proper location of solar power plants. Thus, if the location of new power plants are limited to the SEA-identified REDZ, one should be confident that significant adverse landscape impacts are unlikely to occur from the collective location of these power plants. Subsequently, the risk from cumulative impacts outside of the REDZ should theoretically be higher as such areas were deemed unsuitable to ensure development without significant adverse impacts through the SEA process. In addition, and in contrast to EIAs which are a legislative requirement, the findings of a SEA are primarily used to guide development and expedite environmental authorisations in predetermined areas. According to Therivel[49], the ultimate aim of a SEA is 'to help protect the environment and promote sustainability'. However, considering that only 15% of PV projects are located within the REDZ, alongside the three points of feedback on SEA, throws into question the effectiveness of a SEA in fulfilling this aim. Furthermore, interviewees confirmed that an explanation is needed as to why CSP was not included in the SEA.

The total portion of affected area in the Nama-Karoo and Savanna Biomes under current REIPPPP projects is relatively low and was found to remain low even under a four-fold increase of solar generation capacity. Guided by the solar power capacity allocation in the IRP Update Base Case scenario, an approximated combined area representing 1.88% of the Nama-Karoo and Savanna Biomes is expected to be under solar power development by 2030, according to known projections. However, refinement is needed on the assumptions that were used for this analysis for a more reliable estimate. Similar projections and calculations on expected future solar power development would also assist with the consideration of trade-offs of renewable energy resources as a provisioning ecosystem service compared to the impact of such developments on other supporting ecosystem services.[50]

Conclusion

In light of available energy plans and policy documents, solar power developments are expected to increase in the arid biomes of South Africa. The increase of these developments in these arid areas provides a novel research field of which the findings can be used to inform future development. The relatively small footprints attributed to the two studied solar power technologies deployed in the respective biomes and the current EIA process seem to limit the severity of impacts as experienced and measured at the time of investigation. In addition, the most significant impacts appear to be associated with the construction stage of a development, which represents approximately 10% of the lifespan of a solar power plant. Experience from existing power plants suggests that certain impacts remain excluded from the EIA process, of which cumulative impacts are a key concern and need to be addressed. All aspects considered, a key recommendation is the collection and dissemination of impact monitoring data at multiple solar power plants to feed back into strategic planning for future project siting, which would increase knowledge of solar power development related impacts. This initial exploration provides several potential starting points for the collection of such data and further studies in the arid biomes.

Acknowledgements

We acknowledge financial support from the Centre of Renewable and Sustainable Energy Studies of Stellenbosch University and the National Research Foundation (South Africa), without which this study would not have been possible. We also are grateful for the valuable insights, contribution and time offered to this study by the interviewees and site visit hosts.

Authors' contributions

J.R.: study/project design; data collection; data analyses; conceptualisation and compilation of manuscript. P.G.: study/project supervision; assistance with data presentation; manuscript draft revision. K.J.E.: study/project supervision; manuscript draft revision.

References

1. Scholvin S. South Africa's energy policy: Constrained by nature and path dependency. J S Afr Stud. 2014;40(1):185–202. http://dx.doi.org/10.1080/03057070.2014.889361

2. South African Department of Energy. Integrated resource plan for electricity 2010–2030 [document on the Internet]. c2011 [cited 2016 Dec 05]. Available from: http://www.energy.gov.za/IRP/irp files/IRP2010_2030_Final_Report_20110325.pdf

3. South African Department of Energy. Draft integrated resource plan for electricity 2010: Update [document on the Internet]. c2013 [cited 2016 Dec 05]. Available from: http://www.doe-irp.co.za/content/IRP2010_updatea.pdf

4. South African Department of Environmental Affairs. South African Renewable Energy EIA Application Database [database on the Internet]. c2016 [cited 2016 Mar 10]. Available from: http://egis.environment.gov.za/

5. National Environmental Management Act, Act no. 107 of 1998, Republic of South Africa. Available from: https://www.environment.gov.za/sites/default/files/legislations/nema_impactassessment_g33306rg9314gon543.pdf

6. Burkhardt JJ 3rd, Heath GA, Turchi CS. Life cycle assessment of a parabolic trough concentrating solar power plant and the impacts of key design alternatives. Environ Sci Technol. 2011;45(6):2457–2464. http://dx.doi.org/10.1021/es1033266

7. Hosenuzzaman M, Rahim NA, Selvaraj J, Hasanuzzaman M, Malek ABMA, Nahar A. Global prospects, progress, policies, and environmental impact of solar photovoltaic power generation. Renew Sustain Energy Rev. 2015;41:284–297. http://dx.doi.org/10.1016/j.rser.2014.08.046

8. Lovich JE, Ennen JR. Wildlife conservation and solar energy development in the desert southwest, United States. Bioscience. 2011;61(12):982–992. http://dx.doi.org/10.1525/bio.2011.61.12.8

9. Horner RM, Clark CE. Characterizing variability and reducing uncertainty in estimates of solar land use energy intensity. Renew Sustain Energy Rev. 2013;23:129–137. http://dx.doi.org/10.1016/j.rser.2013.01.014

10. Tsoutsos T, Frantzeskaki N, Gekas V. Environmental impacts from the solar energy technologies. Energy Policy. 2005;33(3):289–296. http://dx.doi.org/10.1016/s0301-4215(03)00241-6

11. Ravi S, Lobell DB, Field CB. Tradeoffs and synergies between biofuel production and large solar infrastructure in deserts. Environ Sci Technol. 2014;48(5):3021–3030. http://dx.doi.org/10.1021/es404950n

12. Turney D, Fthenakis V. Environmental impacts from the installation and operation of large-scale solar power plants. Renew Sustain Energy Rev. 2011;15(6):3261–3270. http://dx.doi.org/10.1016/j.rser.2011.04.023

13. Hernandez RR, Easter SB, Murphy-Mariscal ML, Maestre FT, Tavassoli M, Allen EB, et al. Environmental impacts of utility-scale solar energy. Renew Sustain Energy Rev. 2014;29:766–779. http://dx.doi.org/10.1016/j.rser.2013.08.041

14. McCrary MD, McKernan RL, Schreiber RW, Wagner WD, Sciarrotta TC. Avian mortality at a solar energy power plant. J Field Ornithol. 1986;57(2):135–141.

15. Kagan R, Viner T, Trail P, Espinoza E. Avian mortality at solar energy facilities in southern California: A preliminary analysis [document on the Internet]. c2014 [cited 2016 Jul 12]. Available from: http://alternativeenergy.procon.org/sourcefiles/avian-mortality-solar-energy-ivanpah-apr-2014.pdf

16. Council for Scientific and Industrial Research. Strategic environmental assessment for wind and solar PV energy in South Africa – renewable energy development zones (REDZs) [document on the Internet]. c2014 [cited 2016 Aug 22]. Available from: https://redzs.csir.co.za/

17. Milton SJ, Dean WRJ. Repairing compound damage in arid ecosystems – challenges and controversies. Trans R Soc S Afr. 2015;70(2):127–133. http://dx.doi.org/10.1080/0035919x.2015.1046971

18. Rudman J. Investigating the direct environmental impacts of emerging solar power and shale gas developments in two arid biomes of South Africa [thesis]. Stellenbosch: Stellenbosch University; 2017.

19. Bryman A. Social research methods. 5th ed. Oxford: Oxford University Press; 2015.

20. Noy C. Sampling knowledge: The hermeneutics of snowball sampling in qualitative research. Int J Soc Res Methodol. 2008;11(4):327–344. http://dx.doi.org/10.1080/13645570701401305

21. Friese S. ATLAS. ti 7 User guide and reference [document on the Internet]. c2014 [cited 2016 Mar 10]. Available from: https://www.researchgate.net/publication/264158353_ATLASti_7_User_guide_and_reference

22. Gibbs GR. Thematic coding and categorizing. In:Gibbs GR, editor. Qualitative research kit: Analyzing qualitative data. London: SAGE; 2007. p. 28–55. Available from: http://methods.sagepub.com/book/analyzing-qualitative-data/n4.xml

23. Joffe H, Yardley L. Content and thematic analysis. In: Marks DF, Yardley L, editors. Research methods for clinical and health psychology. Thousand Oaks, CA: SAGE; 2004. p. 56–68.

24. Buskirk T. Significance level. In: Lavrakas P, editor. Encyclopedia of survey research methods. Thousand Oaks, CA: SAGE; 2008. p. 820.

25. Lavrakas P. p-Value. In: Lavrakas P, editor. Encyclopedia of survey research methods. Thousand Oaks, CA: SAGE; 2008. p.648.

26. Addinsoft. XLSTAT 2015 Help documentation [document on the Internet]. c2015 [cited 2016 Mar 10]. Available from: http://assistly-production.s3.amazonaws.com/211808/kb_article_attachments/71753/xlstat_help_EN_original.pdf?AWSAccessKeyId=AKIAJNSFWOZ6ZS23BMKQ&Expires=1474008646&Signature=DKBh6gCFVUESRwHY+JYVySua/h4=&response-content-disposition=filename='xlst

27. Babbie E. The practice of social research. 12th ed. Belmont, CA: Cengage; 2010.

28. Picardi CA, Masick KD. Research methods: Designing and constructing research methods with a real-world focus. Thousand Oaks, CA: SAGE; 2014.

29. GeoTerra Image. 2013-2014 GTI SA National Land-Cover [data set on the Internet]. c2014 [cited 2017 Jul 10]. Available from: http://bgis.sanbi.org/SpatialDataset/Detail/496

30. Mucina L, Rutherford MC, editors. The vegetation of South Africa. Pretoria: South African National Biodiversity Institute; 2006.

31. South African National Parks. National Protected Areas Expansion Strategy [data set on the Internet]. c2010 [cited 2016 Mar 10]. Available from: http://bgis.sanbi.org/SpatialDataset

32. BirdLife South Africa. Important Bird Areas [homepage on the Internet]. c2015 [cited 2016 Aug 11]. Available from: http://www.birdlife.org.za/conservation/important-bird-areas

33. Council for Scientific and Industrial Research. DEA National Strategic Environmental Assessment for the efficient and effective rollout of wind and solar photovoltaic energy [homepage on the Internet]. c2013 [cited 2016 Mar 10]. Available from: http://www.csir.co.za/nationalwindsolarsea/background.html

34. South African Department of Energy. Renewable Energy IPP Procurement Programme (REIPPPP), Bid Window 4 Preferred Bidders Announcement [homepage on the Internet]. c2015 [cited 2016 Aug 22]. Available from: http://www.ipprenewables.co.za/page/304#page/2183

35. Ek K. Public and private attitudes towards 'green' electricity: The case of Swedish wind power. Energy Policy. 2005;33(13):1677–1689. http://dx.doi.org/10.1016/j.enpol.2004.02.005

36. Tsantopoulos G, Arabatzis G, Tampakis S. Public attitudes towards photovoltaic developments: Case study from Greece. Energy Policy. 2014;71:91–106. http://dx.doi.org/10.1016/j.enpol.2014.03.025

37. Karlstrom H, Ryghaug M. Public attitudes towards renewable energy technologies in Norway. The role of party preferences. Energy Policy. 2014;67:656–663. http://dx.doi.org/10.1016/j.enpol.2013.11.049

38. Hernandez RR, Hoffacker MK, Field CB. Land-use efficiency of big solar. Environ Sci Technol. 2014;48:1315–1323. http://dx.doi.org/10.1021/es4043726

39. Hernandez RR, Hoffacker MK, Field CB. Efficient use of land to meet sustainable energy needs. Nat Clim Chang. 2015;(5):353–358. http://dx.doi.org/10.1038/nclimate2556

40. Fluri TP. The potential of concentrating solar power in South Africa. Energy Policy. 2009;37(12):5075–5080. http://dx.doi.org/10.1016/j.enpol.2009.07.017

41. South African Department of Water Affairs and Forestry. National Water Resource Strategy [document on the Internet]. c2004 [cited 2016 Mar 10]. Available from: https://www.dwa.gov.za/documents/Other/StrategicPlan/NWRS2-Final-email-version.pdf

42. Dwernychuk L, Boag D. Ducks nesting in association with gulls – an ecological trap? Can J Zool. 1972;50(5):559–563. http://dx.doi.org/10.1139/z72-076

43. Smith HA. Guidelines to minimise the impact on birds of solar facilities and associated infrastructure in South Africa [homepage on the Internet]. c2012 [cited 2016 Jul 11]. Available from: http://www.the-eis.com/data/literature/Solarguidelines_version2.pdf

44. Reyers B, Rouget M, Jonas Z, Cowling RM, Driver A, Maze K, et al. Developing products for conservation decision-making: Lessons from a spatial biodiversity assessment for South Africa. Divers Distrib. 2007;13(5):608–619. http://dx.doi.org/10.1111/j.1472-4642.2007.00379.x

45. Mace GM, Norris K, Fitter AH. Biodiversity and ecosystem services: A multilayered relationship. Trends Ecol Evol. 2012;27(1):19–26. http://dx.doi.org/10.1016/j.tree.2011.08.006

46. Schmitz OJ. Ecology and ecosystem conservation. Washington DC: Island Press; 2007.

47. Andrews A. Fragmentation of habitat by roads and utility corridors: A review. Aust Zool. 1990;26(3):130–141.

48. Hernandez R, Hoffacker M, Murphy-Mariscal M, Wu G, Allen H. Solar energy development impacts on land cover change and protected areas. Proc Natl Acad Sci USA. 2015;112(44):13579–13584. http://dx.doi.org/10.1073/pnas.1517656112

49. Therivel R. Strategic environmental assessment in action. 2nd ed. New York: Routledge; 2012.

50. Reyers B, O'Farrel P, Cowling R, Egoh B, Le Maitre D, Vlok J. Ecosystem services, land-cover change, and stakeholders: Finding a sustainable foothold for a semiarid biodiversity hotspot. Ecol Soc. 2009;14(1):23.

Student throughput variables and properties: Varying cohort sizes

AUTHOR:
Lucas C.A. Stoop[1]

AFFILIATION:
[1]Independent Researcher, Johannesburg, South Africa

CORRESPONDENCE TO:
Lucas Stoop

EMAIL:
Lcastoop@gmail.com

DATES:
Received: 19 Jan. 2017
Revised: 13 June 2017
Accepted: 28 Aug. 2017

KEYWORDS:
throughput efficiency; student success rates; analytical process model; graduates; dropouts

HOW TO CITE:
Stoop LCA. Student throughput variables and properties: Varying cohort sizes. S Afr J Sci. 2017;113(11/12), Art. #2017-0019, 3 pages. http://dx.doi.org/10.17159/sajs.2017/20170019

ARTICLE INCLUDES:
× Supplementary material
× Data set

FUNDING:
None

A recent research paper described how student throughput variables and properties combine to explain the behaviour of stationary or simplified throughput systems. Such behaviour can be understood in terms of the locus of a point in the triangular admissible region of the H-S plane, where H represents headcounts and S successful credits, each depending on the system properties at that point. The efficiency of the student throughput process is given by the ratio S/H. Simplified throughput systems are characterised by stationary graduation and dropout patterns of students as well as by annual intakes of student cohorts of equal size. The effect of varying the size of the annual intakes of student cohorts is reported on here. The observations made lead to the establishment of a more generalised student throughput theory which includes the simplified theory as a special case. The generalised theory still retains the notion of a triangular admissible region in the H-S plane but with the size and shape of the triangle depending on the size of the student cohorts. The ratio S/H again emerges as the process efficiency measure for throughput systems in general with unchanged roles assigned to important system properties. This theory provides for a more fundamental understanding of student throughput systems encountered in real life.

Significance:

- A generalised stationary student throughput theory through varying cohort sizes allows for a far better understanding of real student throughput systems.

Introduction

A recent research paper[1] (hereafter referred to as the main paper) described how throughput variables combined with throughput properties determine the behaviour of stationary student throughput systems. Throughput variables refer to headcounts (H) and successful credits (S), and throughput properties relate to the percentages of student intakes graduating (G) or dropping out (D) from the degree annually. These percentages also determine on average the number of years for students to graduate (J) or to drop out of a degree (K). Simplified throughput system behaviour can be described by the locus of a point of H and S values in a two-dimensional triangular region, with each admissible point associated with specific system properties. Throughput process efficiency is defined by the ratio S/H. Simplified throughput systems are characterised by stationary graduation and dropout patterns of students – the consequence of annual intakes of student cohorts of equal size.

The effect of varying the size of the annual intakes of student cohorts in student throughput systems for 4-year degrees is reported on here. The assumption of annual intakes of student cohorts of equal size is important in the development of a simplified theory on student throughput, but unfortunately also a constraint when applying the simplified theory to real throughput systems in which annual student intake cohorts vary in size from year to year. A more generalised student throughput theory is therefore required to gain a more fundamental understanding of throughput systems encountered in real life. The development of such a generalised theory is reported on here and it is demonstrated that the simplified theory is indeed a special case of a more generalised theory. The focus will again be on 4-year degrees but the theory as presented can easily be extended to apply to 2- and 3-year degrees.

Generalised cohort survival model calculations

In the case of a 4-year degree with 8 years assumed to be the maximum time for the completion of the degree, cohort studies would require a set of eight consecutive annual intakes N^i (i=1,…,8) of different student cohort sizes. According to the convention followed in the main paper, Cohort 1 is taken to be the youngest cohort of students to be enrolled and Cohort 8 the oldest. The throughput profile of each cohort is firstly defined by the percentage G_j of an intake graduating after j years of study. The throughput profile is secondly defined by the percentage D_j of an intake dropping out from the system after j years of study (part of a year is assumed to be a full year). The following equations will then apply:

$$\left.\begin{aligned} G &= (G_4 + \ldots + G_8); \\ D &= (D_1 + \ldots + D_8) = 1 - G; \text{ and} \\ N &= (N^1 + \ldots + N^8)/8 \end{aligned}\right\} \quad \text{Equation 1}$$

where N is the average cohort size of the throughput system.

On average, the number of years taken by students to graduate at the end of the year (ranging between 4 and 8) is given by:

$$J = (4G_4 N^4 + \ldots + 8G_8 N^8) / (G_4 N^4 + \ldots + G_8 N^8) \quad \text{Equation 2}$$

It is noted that lines of constant J are defined by different G_j values all being proportional to one another. The number of years on average taken by students dropping out at the end of the year (ranging between 1 and 8) is given by:

$$K = (1D_1 N^1 + \ldots + 8D_8 N^8) / (D_1 N^1 + \ldots + D_8 N^8) \quad \text{Equation 3}$$

The headcount of students eventually graduating can be calculated as:

$$H_G = G_4 (N^1 + \ldots + N^4) + G_5 (N^1 + \ldots + N^5) + \ldots$$
$$+ G_7 (N^1 + \ldots + N^7) + G_8 (N^1 + \ldots + N^8) \quad \text{Equation 4}$$

and for those students who will eventually drop out of the degree, the headcount is calculated as:

$$H_D = D_1 (N^1) + D_2 (N^1 + N^2) + D_3 (N^1 + \ldots + N^3) + \ldots$$
$$+ D_7 (N^1 + \ldots + N^7) + D_8 (N^1 + \ldots + N^8) \quad \text{Equation 5}$$

The total headcount as a system variable is therefore defined in terms of the independent sets of system properties G_j and D_j for a given set of N^i as:

$$H = H_G + H_D \quad \text{Equation 6}$$

The annual number of successful module credits earned by students who will eventually graduate can be calculated as:

$$S_G = (4/4)G_4 (N^1 + \ldots + N^4) + (4/5)G_5 (N^1 + \ldots + N^5) + \ldots$$
$$+ (4/7)G_7 (N^1 + \ldots + N^7) + (4/8)G_8 (N^1 + \ldots + N^8) \quad \text{Equation 7}$$

and for students eventually dropping out of the degree, the annual successful module credits are given by:

$$S_D = C H_D = 0.25 H_D \quad \text{Equation 8}$$

where C=0.25 is the number of successful credits earned on average by each of these students per year as is explained in the main paper. The total successful module credits earned annually by all students is therefore defined in terms of the independent sets of system properties G_j and D_j for a given set of N^i as:

$$S = S_G + S_D \quad \text{Equation 9}$$

Relationship between H and S in the generalised throughput system

In the main paper it was demonstrated that for stationary throughput systems, only those combinations of H and S belonging to the admissible region of the H-S plane can be realised. As shown in Figure 1 of the main paper, the admissible region is a triangle WXZ which is bounded by the line WX (J=4), the line WZ (J=8), and the horizontal lines of G=0% and G=100%. The parameter K was set equal to 4, and this value also defines the position of the pivot W of lines of constant J and their intersection with the line G=0%. The perfect throughput system is located at X defined by H=4N=S. In triangle WXZ for a specific value of K, each combination of H and S will correspond to a unique combination of G and J, and vice versa.

In generalising these results, a distinction will now be made between student throughput systems increasing in size as defined by:

$$(N^1 + N^2 + N^3 + N^4) > 4N \quad \text{Equation 10}$$

and student throughput systems decreasing in size as defined by:

$$(N^1 + N^2 + N^3 + N^4) < 4N \quad \text{Equation 11}$$

Student throughput systems where

$$(N^1 + N^2 + N^3 + N^4) = 4N = (N^5 + N^6 + N^7 + N^8) \quad \text{Equation 12}$$

will be referred to as constant size throughput systems. The latter include stationary throughput systems as described in the main paper where all cohorts are of equal size.

Figure 1 demonstrates how the shape of the triangular admissible region for stationary throughput systems with K=4 (solid line triangle) is skewed in the case of student throughput systems increasing in size (triangle WXZ for compound growth of 10% pa) and in the case of student throughput systems decreasing in size (triangle wxZ for compound growth of -10% pa). In triangle WXZ, for example, the coordinates of the vertices of the triangle (in units of N) become:

Vertex X where $G = G_4 = 100\%$:

$$H/N = G_4 (N^1 + N^2 + N^3 + N^4)/N = (N^1 + N^2 + N^3 + N^4)/N = S/N \quad \text{Equation 13}$$

Vertex W where $D = D_4 = 100\%$ and $G = 0\%$:

$$H/N = D_4 (N^1 + N^2 + N^3 + N^4)/N = (N^1 + N^2 + N^3 + N^4)/N = 4S/N \quad \text{Equation 14}$$

Vertex Z where $G = G_8 = 100\%$:

$$H/N = G_8 (N^1 + \ldots + N^8)/N = 8 = 2S/N \quad \text{Equation 15}$$

The triangle WXZ is again bounded by the line WX (J=4) along which $G_4<100\%$ and $G_5=G_6=G_7=G_8=0\%$. The admissible region is also bounded by the line WZ (J=8) along which $G_8<100\%$ and $G_4=G_5=G_6=G_7=0\%$. The boundary XZ represents the line $G_4+G_5+G_6+G_7+G_8=100\%$ which in general is an irregular line shaped by the specific values of N^i.

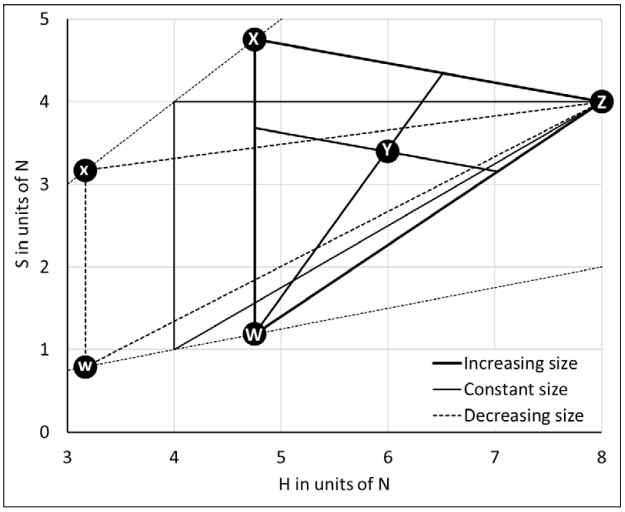


Figure 1: Admissible regions for H (headcounts) and S (successful credits) in the H-S plane for student throughput systems increasing or decreasing in size, as well as constant size systems.

It is interesting to note that the vertex X of all student throughput systems, whether increasing or decreasing in size, all lie on the line S/H=1 which is the condition defining a perfect throughput system. Hence, this condition not only applies to stationary throughput systems, but also to student throughput systems of increasing and decreasing size in general, which emphasises the importance of the condition S/H=1 as a general measure of student throughput performance. Likewise, the vertex W of all student throughput systems, whether increasing or decreasing in size, all lie on the line S/H=0.25.

The throughput configuration Y in the triangle WXZ with the values H/N=6.0 and S/N=3.4 is produced by G=70% with $(G_4;G_5;G_6;G_7;G_8)$ = (3%;15%;28%;19%;5%) for K=4 and a given set of N^i values. As in the case of the simplified throughput model, each combination of G_i values corresponds to a specific value of J, which in this case is equal to J=6.0. Lines of constant G and constant J are also shown in Figure 1.

Finally, the number of graduates produced ($G_4N^4+\ldots+G_8N^8$) by the throughput system, is also a throughput system variable to be measured in units of N.

Conclusions

The present study has shown that the generalised theory on student throughput systems provides for a more fundamental understanding of real throughput systems with the simplified theory indeed a special case of the generalised theory. More specifically, the behaviour of student throughput systems depends on whether these systems are increasing, decreasing, or remain constant in size. The total size of the youngest four student cohorts N^1, N^2, N^3 and N^4 relative to the average size of all student cohorts, determines the size and shape of the admissible triangle of H and S values in the H-S plane. One of the advantages of the generalised theory is that the position of the relevant admissible domain in the H-S plane can be positioned accurately, instead of relying on assumptions about constant annual intakes of students. Otherwise, the main characteristics of the generalised student throughput theory appear to be logical extensions of the simplified theory. The ratio S/H again emerges as a measure of throughput process efficiency with unchanged roles assigned to the generalised values of J, K and G.

Regarding the practical use of the generalised theory, a few examples have already been discussed in the main paper to show how the simplified theory could find application to advance our understanding of student throughput systems. The generalised theory certainly would add a further dimension to our understanding of such systems, and would permit the analysis of even more complex student throughput systems which could not have been analysed before.

Reference

1. Stoop LCA. Relationships between student throughput variables and properties. S Afr J Sci. 2015;111(7/8), Art. #2014-0336, 5 pages. http://dx.doi.org/10.17159/sajs.2015/20140336

The influence of collaboration in research priorities: The SADC Case

AUTHOR:
Anastassios Pouris[1]

AFFILIATION:
[1]Institute for Technological Innovation, University of Pretoria, Pretoria, South Africa

CORRESPONDENCE TO:
Anastassios Pouris

EMAIL:
apouris@icon.co.za

DATES:
Received: 18 May 2017
Revised: 12 July 2017
Accepted: 12 Sep. 2017

KEYWORDS:
impact; scientifically small countries; South Africa; research collaboration; co-authorship

HOW TO CITE:
Pouris A. The influence of collaboration in research priorities: The SADC case. S Afr J Sci. 2017;113(11/12), Art. #2017-0150, 3 pages. http://dx.doi.org/10.17159/sajs.2017/20170150

ARTICLE INCLUDES:
× Supplementary material
× Data set

FUNDING:
National Research Foundation (South Africa)

This study was aimed at providing evidence of the effects of collaboration among unequal partners on their research priorities. Co-authorship patterns were investigated among South African authors publishing with authors of other countries in the region, with and without other non-African co-authors. It was identified that the non-African collaborators have a high impact on the quantity of co-authored publications and on the research disciplines in which co-authored research is undertaken. The findings raise a number of policy questions.

Significance:

- The findings make profound that African countries should prioritise and engage their limited resources in areas of national priorities.

Introduction

Research collaboration is on an ascending path and receives particular attention from governments and researchers internationally.[1] Researchers investigate the modes, effects and motives of collaborating researchers, while governments attempt to utilise research collaboration as an instrument for technology transfer from universities and science councils to industry (intra-collaboration); for know-how transfer from abroad (inter-collaboration); and as a means of improving diplomatic relations with other governments by creating goodwill and political capital[2]; among other reasons.

The literature identifies that researchers collaborate with each other for a number of reasons. Reasons for collaboration include improving their visibility and recognition[3], utilising expensive equipment and facilities which are not under their control[4]; and acquiring expertise and new ideas needed for their research[5]. Other reasons include historical ties; linguistic preferences; geographical proximity; and specific problem issues (e.g. disease control or natural disaster mitigation).

In the domain of policy, research collaboration has become an important component of science, technology and innovation internationally and substantial resources are allocated by governments (e.g. South Africa, the European Commission and the USA) for this objective.

There are several articles that report on the investigation of collaboration on the African continent. Sooryamoorthy[6] investigated the collaboration patterns of researchers in South Africa and Boshoff[7] investigated the collaborative patterns of authors in the Southern African Development Community (SADC). Boshoff identified that 'only 3% of SADC papers during 2005–2008 were jointly authored by researchers from two or more SADC countries (intra-regional collaboration), and only 5% of SADC papers were jointly authored with researchers from African countries outside the SADC (continental collaboration)'[7(p.481)]. Similarly, Onyancha and Maluleka[8] identified that knowledge production through co-authored research among researchers from sub-Saharan African countries is minimal.

More recently, Pouris and Ho[9] identified the status of research co-authorship on the African continent during the period 2007–2011. The results are based on the analysis of more than 111 000 articles which had at least one author with a corporate address within the African continent. Estimation of the activity indices of various scientific fields shows the emphasis or under emphasis of the various fields. The activity index is defined as a country's share in the world's publication output in the particular field divided by the country or region's share in the world's publication output in all disciplinary fields.

The most emphasised scientific disciplines are: Tropical Medicine (12.5 times higher than what should be expected from the research size of Africa); Parasitology (6.5 times higher) and Infectious Diseases (4.6 times higher). It is apparent that the emphasised research areas are dominated by medical and natural resources fields (Biodiversity; Water Resources; Entomology; Mining, etc.). Furthermore, the individual African countries were identified to exhibit higher collaboration patterns than countries of other continents. Nigeria was estimated to be the only country with a co-authorship rate lower than 50%. A total of 29 African countries were identified to publish more than 90% of their articles in collaboration with authors from other countries. The authors argued that this pattern is indicative of dependency on foreign resources.[9]

The main countries co-authoring research with Africa were identified as the USA, France and the UK. The authors emphasise that these countries are the most collaborative countries in the world.[9] These three countries – USA, France and the UK – are also the largest funders of research in biosciences, with more emphasis on medicine and agricultural sciences, in Africa. Furthermore, collaboration with non-African countries exceeds that of inter-African collaboration. The authors state: 'It is logical to argue that African collaboration is not driven by local researchers searching for collaborators, but by the availability of resources and interests outside the continent.' As an anonymous reviewer stated: 'Non-African research funding often steers how African scientists choose research partners and topics.'

The above is of particular importance as they indicate that research conditions on the continent are amenable to direction from outside interests.

Pouris[10] investigated the research performance of the 15 countries in the SADC region. It was estimated that South Africa – with 19% of the population of the region – is responsible for 60% of the GDP in the region and produces 79% of the region's publications. All SADC countries appear to have the same focus in their research priorities and underemphasise disciplines such as engineering, materials science and molecular biology. The author[10] expressed concern that the current research structures are inadequate to assist in reaching the objectives developed in the Regional Indicative Strategic Development Plan of the Community.

Zdravkovic et al.[11], through interviews in four SADC countries, identified that the interviewed scientists working with scientists in the North enjoyed better funding, more organised research and access to different knowledge than those without Northern collaboration. However, South–South collaboration meant easier contact, working under equal conditions, and solving relevant problems for Africa.

Sooryamoorthy[12] investigated the interplay of collaboration and citations in the social sciences in South Africa and other authors[13,14] have examined collaboration in particular scientific disciplines.

In summary, the majority of the African-related investigations have been focused on the effects of collaborations on impact, as it is manifested in citations, and the meagre size of inter-African collaboration. However, a critical issue in terms of policy is the possible effects of collaboration on the priorities and direction of the innovation system. This study aimed to identify whether international collaborators influence the size and research disciplines in the South African–SADC collaboration. The issue is of particular importance when scientifically large countries collaborate with scientifically small countries. This short communication further aims to make a contribution in this domain investigating collaboration and priorities in the SADC region.

Methodology

The scientometric analysis was facilitated by Clarivate Analytics' (formerly Thomson Reuters) Web of Science databases. The Web of Science covers comprehensively the most important journals in the world and most importantly provides information about all authors in an article. The latter is particularly advantageous when the objective of the effort is to investigate collaborative efforts.

The database was interrogated in order to identify the collaborative patterns of South African researchers. All publications with at least one South African address were identified and were analysed. Using the field tag 'country' we identified the SADC countries collaborating with South Africa. Subsequently, the collaborative disciplines between South Africa and the other SADC countries were identified (using the field tag 'research areas'). This list of scientific disciplines was compared with the same list after excluding all articles with non-African collaborators.

The 3-year period 2012–2014 was chosen to be investigated as it is the period after the completion of the 7th Framework Programme (FP7). FP7 was a powerful European instrument which promoted collaborative research and development (among others).

Results

During the 2012–2014 period, South African researchers produced 45 343 publications. The majority of these publications were articles (85.7%). Other outputs were meeting abstracts (79); review articles (64); book chapters (51); editorial material (38); proceedings papers (27) and others.

Of the research articles, 23 581 (52%) were co-authored with at least one author from another country. Among the co-authored publications 1505 (6.4%) had at least one co-author from the SADC region. The main collaborating countries were Zimbabwe (406 articles), Malawi (237 articles) and Namibia (221).

In order to identify the influence of non-African countries in the regional co-authorship effort, the articles that had non-African co-authors were excluded. Hence, only 563 publications were identified to be co-authored between South African and SADC co-authors (without non-African co-authors) – only 2.4% of the South African co-authored publications.

Table 1 shows the research areas in which South Africa and SADC countries collaborate with all collaborators. It becomes apparent that medical and health issues dominate the co-authorship list.

Table 2 shows the most prolific research areas in the cooperation efforts when there are no non-African participants. Agriculture and Environmental Sciences Ecology are top of the list. It is interesting to note that Infectious Diseases and Immunology that were on top of the list in Table 1, fall lower in the list of Table 2. When there is no non-African influence, the co-authorship priorities appear to change. Infectious Diseases and Immunology appear to be led by foreign researchers. It is emphasised that even though these figures are relatively small, they are the total populations of articles with the particular characteristics.

Table 1: Research areas of collaborative research: South African authors with SADC and other co-authors

Research area	Number of publications out of 1505
Infectious Diseases	214 (14.2%)
Immunology	160 (10.6%)
Public Environmental Occupational Health	131 (8.7%)
Environmental Sciences Ecology	130 (8.6%)
Science Technology other topics	109 (7.2%)
Agriculture	68 (4.5%)
Virology	64 (4.2%)
General Internal Medicine	62 (4.1%)
Geology	56 (3.7%)
Microbiology	49 (3.2%)
Plant Sciences	44 (2.9%)
Zoology	44 (2.9%)
Veterinary Sciences	43 (2.8%)
Tropical Medicine	41 (2.7%)
Astronomy Astrophysics	40 (2.6%)
Water Resources	40 (2.6%)
Chemistry	39 (2.5%)

Table 2: Research areas of collaborative research: South African authors with no non-African co-authors, 2012–2014

Research area	Number of publications out of 563
Agriculture	56 (9.9%)
Environmental Sciences Ecology	44 (7.8%)
Public Environmental Occupational Health	31 (5.5%)
Plant Sciences	30 (5.3%)
Mathematics	25 (4.4%)
Engineering	24 (4.2%)
Physics	24 (4.2%)
Water Resources	24 (4.2%)
Chemistry	22 (3.9%)
Geology	22 (3.9%)
Infectious Diseases	22 (3.9%)

Discussion

The aim was to create awareness of the influence of non-African collaborators in the SADC's research priorities. The choice of the groups has been made so that scientifically big countries (non-African countries) collaborate with a relatively small scientific community. The findings are suggestive: the majority (almost two thirds) of South African–SADC collaboration includes non-African participants. While it is difficult to surmise what would have happened if non-African collaborators were not available, it may be argued that these collaborations were initiated by the non-African participants. What is probably more important is the fact that the collaborations with no non-Africans occurred in disciplines different from those in which non-Africans participated. This finding confirms the assertion of other researchers[15] that most African science collaboration flows through international gates.

A number of policy relevant questions can be raised. Are the collaborative disciplines also induced by the non-African participants? If so, are they in the interest of the local regional system of innovation? What would happen if the non-African participants lose interest in the region? How can local collaboration be improved? There are others.

Further research, including surveys and comparisons in other regions in the world, may confirm the validity of the argument that scientifically big countries have the power to dictate priorities in small regions.

Acknowledgements

This study is based on research wholly supported by the National Research Foundation of South Africa (grant no: 95857).

References

1. Yeung YY, Liu TCY, Ng PH. A social network analysis of research collaboration in physics education. Am J Phys. 2005;73(2):145–150. https://doi.org/10.1119/1.1775799
2. Wagner CS, Staheli L, Silberglitt R, Wong A, Kadtke J. Linking effectively: Learning lessons from successful collaboration in science and technology. Santa Monica, CA: RAND; 2002.
3. Narin F, Stevens K, Whitlow E. Scientific cooperation in Europe and the citation of multinationally authored papers. Scientometrics. 1991;21:313–323. https://doi.org/10.1007/BF02093973
4. Schubert T, Sooryamoorthy R. Can the centre-periphery model explain patterns of international scientific collaboration among threshold and industrialised countries? The case of South Africa and Germany. Scientometrics. 2010;83(1):181–203. https://doi.org/10.1007/s11192-009-0074-2
5. DeB Beaver D, Rosen R. Studies in scientific collaboration: Part I. Professional origins of scientific co-authorship. Scientometrics. 1978;1(1):65–84. https://doi.org/10.1007/BF02016840
6. Sooryamoorthy R. Collaboration and publication: How collaborative are scientists in South Africa? Scientometrics. 2009;80(2):419–439. https://doi.org/10.1007/s11192-008-2074-z
7. Boshoff N. South–South research collaboration of countries in the Southern African Development Community (SADC). Scientometrics. 2010;84(2):481–503. https://doi.org/10.1007/s11192-009-0120-0
8. Onyancha OB, Maluleka JR. Knowledge production through collaborative research in sub-Saharan Africa: How much do countries contribute to each other's knowledge output and citation impact? Scientometrics. 2011;87(2):315–336. https://doi.org/10.1007/s11192-010-0330-5
9. Pouris A, Ho Y-S. Research emphasis and collaboration in Africa. Scientometrics. 2014;98:2169–2184. https://doi.org/10.1007/s11192-013-1156-8
10. Pouris A. A scientometric analysis of SADC countries: Science in the tip of Africa. Scientometrics. 2010;85(1):145–154. https://doi.org/10.1007/s11192-010-0260-2
11. Zdravkovic M, Chiwona-Karltun L, Zink E. Experiences and perceptions of South–South and North–South scientific collaboration of mathematicians, physicists and chemists from five southern African universities Scientometrics. 2016;108:717–743. https://doi.org/10.1007/s11192-016-1989-z
12. Sooryamoorthy R. Do types of collaboration change citation? A scientometric analysis of social sciences publications in South Africa. Scientometrics. 2017;111(1):379–400. https://doi.org/10.1007/s11192-017-2265-6
13. Mehdi S, Jenab H. Two dimensional mapping of scientific production of nations in the fields of physics and astronomy. S Afr J Sci. 2016;112(5/6), Art. #2015-0119, 8 pages. https://doi.org/10.17159/sajs.2016/20150119
14. Pouris A, Ho Y-S. A bibliometric analysis of research on Ebola in Science Citation Index Expanded. S Afr J Sci. 2016;112(3/4), Art. #2015-0326, 6 pages. https://doi.org/10.17159/sajs.2016/20150326
15. Toivanen H, Ponomariov B. African regional innovation systems: Bibliometric analysis of research collaboration patterns 2005–2009. Scientometrics. 2011;88(2):471–493. https://doi.org/10.1007/s11192-011-0390-1

Dispersal of semi-fleshy fruits to rock crevices by a rock-restricted rodent

AUTHORS:
Joseph D.M. White[1]
Jeremy J. Midgley[1]

AFFILIATION:
[1]Department of Biological Sciences, University of Cape Town, Cape Town, South Africa

CORRESPONDENCE TO:
Joseph White

EMAIL:
josephwhite16@gmail.com

DATES:
Received: 16 May 2017
Revised: 11 Sep. 2017
Accepted: 12 Oct. 2017

KEYWORDS:
seed dispersal; directed dispersal; *Heeria argentea*; rock outcrops; fire refugia

HOW TO CITE:
White JDM, Midgley JJ. Dispersal of semi-fleshy fruits to rock crevices by a rock-restricted rodent. S Afr J Sci. 2017;113(11/12), Art. #2017-0159, 5 pages. http://dx.doi.org/10.17159/sajs.2017/20170159

ARTICLE INCLUDES:
✓ Supplementary material
× Data set

FUNDING:
National Research Foundation (South Africa)

Seed dispersal allows successive generations of plants to be mobile in space and time. *Heeria argentea*'s unusual fruit and its ubiquity in extremely rocky habitats, suggests that this tree requires a specialist disperser. We therefore investigated the dispersal ecology of *H. argentea* and *Hartogiella schinoides*. We found *M. namaquensis* rapidly removed *H. argentea* and *H. schinoides* fruits, moving them short distances within and between rock outcrops, and consumed only the pericarps. Birds were observed consuming *H. schinoides,* but not *H. argentea* fruits, suggesting *M. namaquensis* is its sole, specialist disperser. Most *H. argentea* seeds (65%) with removed pericarps germinated successfully, while intact fruits did not. We show rock outcrops represent fire refugia, allowing *H. argentea* trees to grow to large sizes, with small stems and a co-occurring, wind-dispersed tree, *Widdringtonia nodiflora* found away from these sites. This rodent–tree mutualism is perhaps the clearest global example of directed dispersal and shows that these endemic trees are highly adapted for survival in the southwestern Cape habitat and are not tropical relicts.

Significance:

- The fruits of rock-restricted Cape trees are directly dispersed by rock rats to rock outcrops. This is the first description of rodent dispersal of fleshy fruits in South Africa.
- This species-specific interaction allows for rapid germination of seeds and protection from frequent fires for adults. This rodent–tree mutualism is perhaps the clearest global example of directed seed dispersal.

Introduction

Seed dispersal allows plants, via their seeds, to be mobile in space and time. The spatio-temporal locations at which seeds arrive have a major influence on the subsequent fitness of offspring, as location is a major determinant of seed and seedling survival. Many small patches of fire-avoiding forest exist in a wider sea of fire-prone shrublands and savanna across South Africa. In the southwestern Cape, these forests are restricted to growing in extremely rocky habitats, such as on cliffs and amongst rock outcrops and screes.[1]

The dynamics of these forests are different to that of the adjacent shrublands (known as fynbos) where post-fire recruitment from long-lived seed banks of non-fleshy fruits dominates.[2] For example, many Proteaceae as well as *Widdringtonia nodiflora* (mountain cypress; Cupressaceae) are serotinous trees with dry, wind-dispersed seeds that grow in the open shrublands. In contrast, forest tree fruits are typically soft and non-dormant. *Heeria argentea* (rockwood) and *Hartogiella schinoides* (spoonwood) (sensu Islam et al.[3]) are important constituents of these forests. *Heeria argentea* has pale, yellow-green, leathery-coated fruits up to 30 mm in diameter with a 3 mm thin, fleshy pericarp surrounding a single, soft, chlorophyllous seed (Figure 1a,b). Similarly, *H. schinoides* seeds are chlorophyllous but the fruit is smaller (diameter = 15 mm) and more visually distinct being a dark, red colour (Figure 1c,d).

Birds are unlikely dispersers of *H. argentea*, because the fruits are unattractive and unrewarding. Marloth[4] suggested that *Procavia capensis* (rock hyrax) is the primary disperser of *H. argentea*. Presumably his main evidence was their overlapping distributions, both being rock-restricted, and the unusual fruit. Although there are no data, others have cited this hypothesis.[5,6] We investigated the dispersal ecology of these fruiting species based on Marloth's hypothesis.

Methods

Data were collected in April and May 2016 at seven sites in the Limietberg Nature Reserve (33.620355°S, 19.106678°E) in the southwestern Cape, South Africa. *H. argentea* fruits were collected directly off trees in April 2016. The fruits ripen and fall to the ground over several months in the austral summer (personal observation). We placed 5–10 *H. argentea* fruits at depots for 3 days, with six depots per site (n_{total} = six sites; n_{total} = 330 fruits); four sites were in rock outcrops and two sites were in adjacent non-rocky proteoid shrublands. To locate moved fruits, each was attached to a reverse-wound bobbin with fast-setting non-toxic glue. At each depot (placed ± 50 m apart), we used a LTL Acorn 6210M remotely activated camera trap (Shenzen LTL Acorn Electronics Co., Shenzen, Guangdong, China) to monitor animal–fruit interactions. Cameras were set to record one photograph, immediately followed by a 30-s video. Each time an animal interacted with a novel fruit or entered/left the field of view it was considered a new event. For *H. schinoides,* we followed a similar, scaled-down experiment, placing out six depots at a single site with 10 fruits per depot. No camera traps were used in this experiment.

Germination trials were conducted on intact and handled *H. argentea* fruits (n=40 of each) over 6 weeks at the University of Cape Town's glasshouse. Fruits were placed in individual trays on the surface of a potting soil medium and watered once a week. We used a chi-squared test in the R programming language[7] to compare overall germination of fruits.

Scale bar = 2 mm

Figure 1: Intact and cross-sectioned fruits of (a,b) *Heeria argentea* and (c,d) *Hartogiella schinoides*, showing the thin, fleshy pericarps and chlorophyllous endosperm/embryo.

We investigated fire impacts on *H. argentea* and *W. nodiflora* individuals 1 year post-fire. We measured the basal diameter (mm) of *H. argentea* and *W. nodiflora* adults in relation to distance within and from the outer edge of rock outcrops (defined here as a collection of rocks ≥2 m^2). We estimated the fire damage by rating the proportion of the canopy dead in *H. argentea* and *W. nodiflora*. All individuals were surveyed within a 0.6-ha area that included both shrublands and rock outcrops. Beta regression[8] in R[7] was used to determine the relationships between tree size, distance to rock outcrops and fire damage. Reported values are mean ± standard deviation.

Results

Rock outcrops

Within only 3 nights, 6.63±3.85 fruits per depot were removed (total = 66%; n=159/240). Of these 159 fruits, 31% were located and collected, and 40% were not retrieved, but were tracked to a minimum distance until the thread either snapped or was dislodged. The remaining 29% were considered lost, likely as a result of thread failure.

Of the 159 moved fruits that left a thread trail, we determined that 90% of these had been moved at least 0.3 m into rock outcrops 3.26±3.77 m from the initial depot. Of these dispersed fruits, 92% had more than 75% of the pericarp removed (Figure 2a,b). None of the chlorophyllous seeds were visibly damaged. Seeds with pericarps did not germinate (n=0/40) whereas seeds with removed pericarps did (n=26/40, 65%; X^2=31.5, d.f.=1, p<0.001; Figure 2c).

Despite observation of their droppings within 50 m of all sites, only one rock hyrax was recorded on camera at one depot and this individual did not interact with the fruits. Instead, *Micaelamys namaquensis* (Namaqua rock rat) was the most frequently viewed small mammal on camera trap observations (n=151) and was the only animal viewed removing fruits in the rock outcrop sites (n=100 events) (Supplementary video 1). No bird visitation was documented. Fruits were typically moved out of the camera field of view but at one site we observed a Namaqua rock rat consuming the pericarp (Supplementary video 2). We also viewed the following carnivores on the camera trap observations: *Genetta tigrina* (Cape genet) and *Galerella pulverulenta* (Cape grey mongoose), neither of which interacted with fruits.

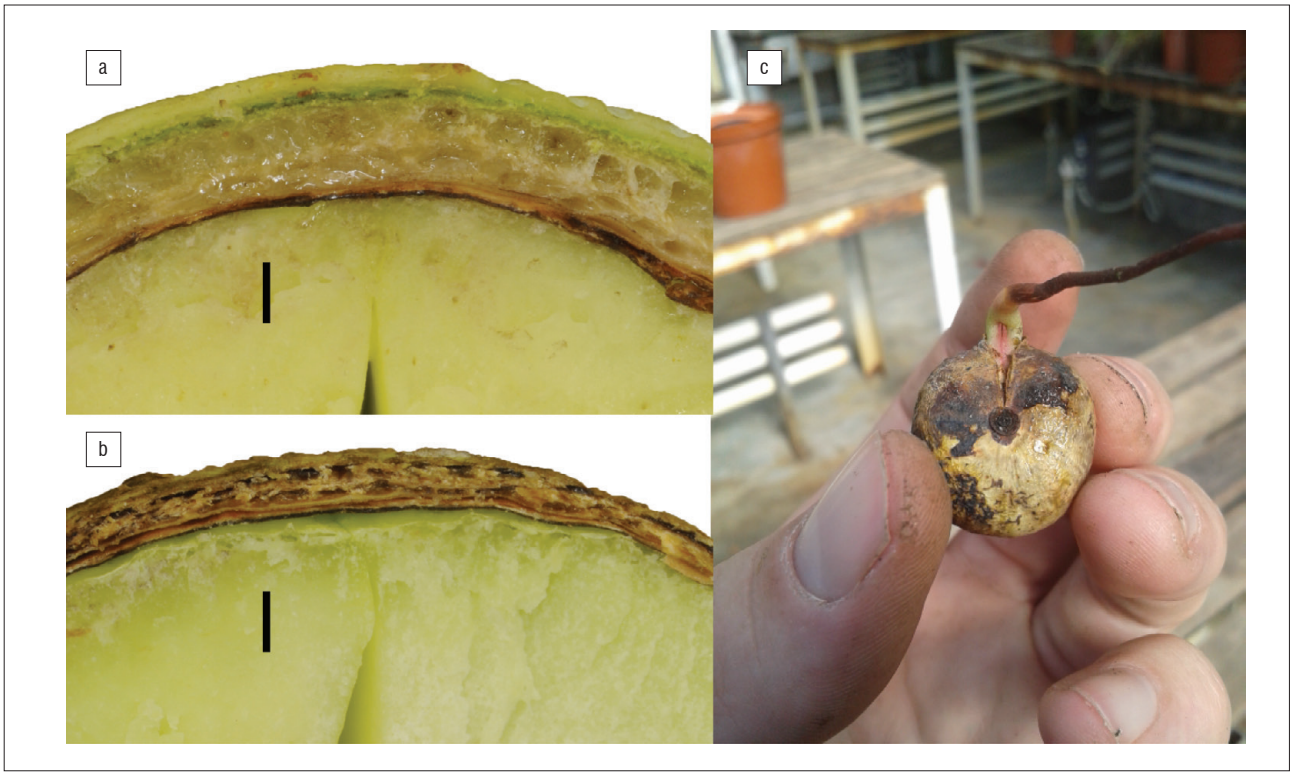
Dispersal of *H. schinoides* appeared to follow the same pattern. Within 3 days, 80% of fruits (n_{total} = 60) were removed and 58% of these were located (n=28), with 93% of those moved to rock outcrops and 68% had their full pericarp removed. We incidentally observed *Onychognathus morio* (red-wing starlings) feeding on *H. schinoides* fruits.

Shrublands

Fruits were rarely moved (2.25±2.31 per depot; 16% of n=90) at shrubland sites, with most located (n=11/14). Fruits occasionally showed the pericarp slightly consumed (n=11), with the pericarp never fully removed. All moved fruits were discarded in the open on the soil surface. Camera trap observations showed that the four small mammal species recorded usually ignored fruits (Supplementary video 3). These species included *Rhadbomys pumilio* (four-striped grass mouse) (n=92), *Otomys irroratus* (vlei rat) (n=46) and *Acomys subspinosus* (Cape spiny mouse) (n=14), as well as occasionally *Hystrix africaeaustralis* (Cape porcupine), with only *R. pumilio* occasionally interested in *H. argentea* fruits (removing n=10 fruits).

Plant distribution and fire damage

Heeria argentea trees were distributed on the margin or within rock outcrops (64%), and 36% were found away from rock outcrops. Large adult *H. argentea* trees were typically found far within rock outcrops (-2.65±2.89 m) and lacked any fire damage. Small *H. argentea* trees were mostly found away from rock outcrops (7.83±8.94 m) and often received full top-kill from fire damage (53% of n=139 trees)(see Figure 3). *W. nodiflora* was never found within rocky refugia and always experienced full top-kill from fire damage. Both basal diameter (estimate ± s.d. = -0.04±0.01, z=-3.57, p<0.001) and distance relative to rock outcrops (0.06±0.01, z=4.58, p<0.001) were significant predictors of fire damage in *H. argentea* (pseudo R^2 = 0.46).



(a,b) Scale bar = 1 mm

Figure 2: Cross-section of *Heeria argentea* fruits that are (a) intact and (b) with the pericarp removed showing the fleshy layer that Namaqua rock rats consume. (c) A germinated seed with the radicle breaking through a fruit without the leathery pericarp.

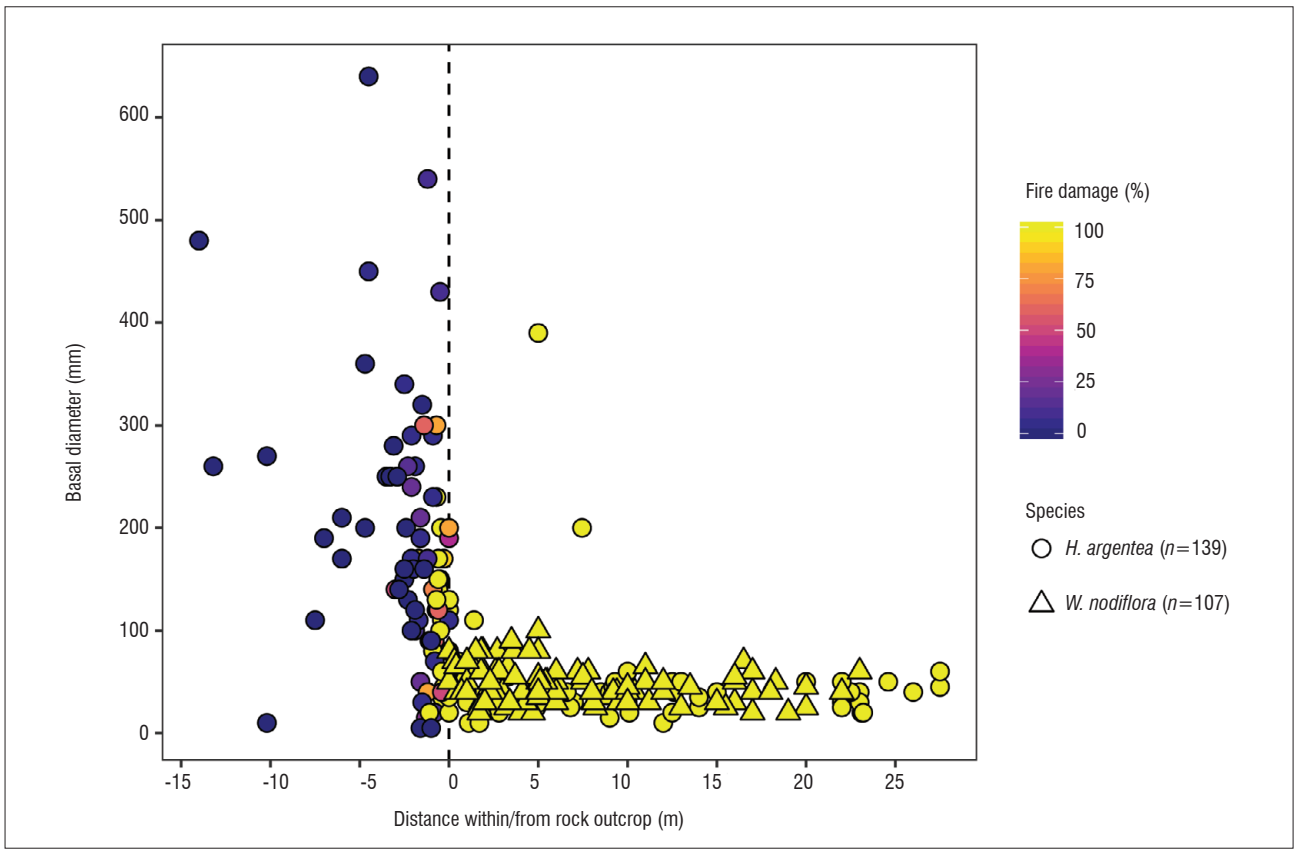


Figure 3: The proportion of canopy damage from a recent fire to *Heeria argentea* and *Widdringtonia nodiflora* trees of different sizes within (negative) and from (positive) rock outcrops.

Discussion

The rapid rate of fruit removal and pericarp consumption showed that the fruits of both *H. argentea* and *H. schinoides* are highly attractive and rewarding to Namaqua rock rats. Our findings thus lead to the rejection of Marloth's[4] original hypothesis. As the seeds have no dormancy, germination is not stimulated by fire, or its products, unlike most seeds in fire-prone ecosystems.[2] *H. argentea* is thus dependent on rock rats for dispersal and rapid initiation of germination. Avoiding predation by carnivores may explain why Namaqua rock rats take fruits to narrow and concealed crevices before consuming the pericarp layer.

The fruit dispersal patterns of *H. schinoides* also suggest Namaqua rock rat dispersal. Their fruits are smaller and dark red in colour – likely an adaptation for bird frugivory. We observed bird frugivory on *H. schinoides* fruits, but never observed birds feeding on *H. argentea*. The fruit of *Heeria* is likely too large for the gape of most local frugivorous birds and, as the seed is soft, pecking, biting or ingestion could destroy it. Further, birds would not likely disperse the seeds into crevices. As Namaqua rock rats also feed on *H. schinoides,* they are likely generalist frugivores.

The significance of dispersal to rock outcrops is not immediately apparent, as these may be poor sites for initial seedling growth. The large seed, and consequent large energy stores, of *H. argentea* may facilitate some growth in deep shade (Figure 4). Large *H. argentea* trees were mostly found within rocky refugia where they experienced almost no fire damage. In contrast, small stems were found away from fire refugia and generally experienced full canopy death. This difference suggests that individuals in less rock-protected sites are condemned to small size and likely reduced reproductive capacity as a result of frequent top-kill and consequent basal resprouting after repeated burning. Wind dispersal of *W. nodiflora* seeds appears to prevent arrival in rocky refugia because no individuals were observed within rock outcrops. Unlike *H. argentea*, all *W. nodiflora* trees experienced full top-kill during the recent fire (Figure 3).

Aspects of the *H. argentea*–Namaqua rock rat interaction are unusual. Firstly, rodent dispersal of fleshy fruits is a novel dispersal mechanism in the Cape. Rodents may be dispersers of non-fleshy, dry seeds typically by scatter-hoarding.[9,10] Namaqua rock rat dispersal of *H. argentea* differs from most other small mammal dispersal. In other rodent frugivory cases, the fruits may be partially damaged.[11-14] Globally, rodent frugivory, in which the seed is not ingested and left undamaged, has only been documented in a few studies.[13,15] Exotic rodents have been documented to disperse seeds while only consuming the pericarp.[13] By consuming only the pericarp, Namaqua rock rats still achieve a reward while leaving the seed intact. Secondly, this is an example of directed dispersal in which fruits disproportionately arrive at non-random, fire-protected destinations that are highly favourable sites for survival.[16]

The three classic examples of directed dispersal are somewhat controversial. Mistletoe seedlings can only establish on thin, sun-exposed branches with only specific bird species, such as *Dicaeum hirundinaceum* (mistletoebirds), considered able to disperse seeds to these sites. That generalist birds, marsupials[17], explosive seed release[18] and the wind[19] also effectively disperse mistletoe seeds weakens this argument. Myrmecochory (ant-dispersal) has been suggested as a mechanism for directed dispersal to ant nests (localised sites of high nutrients in nutrient-poor shrublands), but some evidence suggests otherwise.[20,21] Scatter-hoarding by rodents can be 'directed' to sites of low adult conspecifics.[22] However, the benefit of these sites is escape from seed predators. *H. argentea* fruit dispersal is possibly the clearest global example of directed dispersal – too short-distance, site-specific and beneficial to established plants (rather than seedlings), to be explained by escape of seed or seedling predators or by colonisation of prime sites by seedlings.[16]

Figure 4: A *Heeria argentea* seedling emerges from a dark, rocky crevice at Limietberg Nature Reserve.

Acknowledgements

We thank Deon Rossouw of CapeNature for permission to work in Limietberg Nature Reserve. We also thank the South African National Research Foundation for funding. J.E. Quince provided assistance with data collection.

Authors' contributions

J.J.M. conceptualised the study. Both authors contributed to data collection and writing. J.D.M.W. did the data analyses.

References

1. Moll EJ, McKenzie B, McLachlan D. A possible explanation for the lack of trees in fynbos, Cape Province, South Africa. Biol Conserv. 1980;17:221–228. https://doi.org/10.1016/0006-3207(80)90057-9
2. Le Maitre DC, Midgley JJ. Plant reproductive ecology. In: Cowling R, editor. The ecology of fynbos: Nutrients, fire and diversity. Oxford: Oxford University Press; 1992. p. 135–174.
3. Islam MB, Simmons MP, Archer RH. Phylogeny of the *Elaeodendron* group (Celastraceae) inferred from morphological characters and nuclear and plastid genes. Syst Bot. 2006;31(3):512–524. https://doi.org/10.1600/036364406778388610
4. Marloth R. The flora of South Africa. Vol. II, Section II. Cape Town: Darter Bros. & Co.; 1925.
5. Von Teichman I, Van Wyk AE. Taxonomic significance of pericarp and seed structure in *Heeria argentea* (Thunb.) Meisn. (Anacardiaceae), including reference to pachychalazy and recalcitrance. Bot J Linn Soc. 1996;122:335–352. https://doi.org/10.1111/j.1095-8339.1996.tb02080.x
6. Van Wyk B, Van Wyk P. Field guide to trees of southern Africa. Cape Town: C. Struik; 1997.
7. R Core Team. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2016.
8. Cribari-Neto F, Zeilis A. Beta Regression in R. J Stat Softw. 2010;34:1–24. https://doi.org/10.18637/jss.v034.i02
9. Vander Wall SB. Food hoarding in animals. Chicago, IL: University of Chicago Press; 1990.
10. White JDM, Bronner GN, Midgley JJ. Camera-trapping and seed-labelling reveals widespread granivory and scatter-hoarding of nuts by rodents in the Fynbos Biome. African Zool. 2017;52(1):31–41. https://doi.org/10.1080/15627020.2017.1292861
11. Steele MA, Knowles T, Bridle K, Simms EL. Tannins and partial consumption of acorns: Implications for dispersal of oaks by seed predators. Am Midl Nat. 1993;130(2):229–238. https://doi.org/10.2307/2426123
12. Mendoza E, Dirzo R. Seed tolerance to predation: Evidence from the toxic seeds of the buckeye tree (*Aesculus californica*; Sapindaceae). Am J Bot. 2009;96(7):1255–1261. https://doi.org/10.3732/ajb.0800297
13. Shiels AB, Drake DR. Are introduced rats (*Rattus rattus*) both seed predators and dispersers in Hawaii? Biol Invasions. 2011;13:883–894. https://doi.org/10.1007/s10530-010-9876-7
14. Loayza AP, Carvajal DE, García-Guzmán P, Gutierrez JR, Squeo FA. Seed predation by rodents results in directed dispersal of viable seed fragments of an endangered desert shrub. Ecosphere. 2014;5(4):1–9. https://doi.org/10.1890/ES13-00283.1
15. Feer F, Henry O, Forget P-M, Gayot M. Frugivory and seed dispersal by terrestrial mammals. In: Bongers F, Charles-Dominique P, Forget P, Théry M, editors. Nouragues: Dynamics and plant-animal interactions in a neotropical rainforest. Dordrecht: Springer; 2001. p. 227–232. https://doi.org/10.1007/978-94-015-9821-7_21
16. Wenny DG. Advantages of seed dispersal: A re-evaluation of directed dispersal. Evol Ecol Res. 2001;3:51–74.
17. Amica G, Aizen MA. Mistletoe seed dispersal by a marsupial. Nature. 2000;408:929–930. https://doi.org/10.1038/35050170
18. DeBruyn RAJ, Paetkau M, Ross KA, Godfrey DV, Friedman CR. Thermogenesis-triggered seed dispersal in dwarf mistletoe. Nat Commun. 2015;6(6262):1–5. https://doi.org/10.1038/ncomms7262
19. Tercero-Bucardo N, Kitzberger T. Establishment and life history characteristics of the southern South American mistletoe *Misodendrum punctulatum* (Misodendraceae). Rev Chil Hist Nat. 2004;77:509–521. https://doi.org/10.4067/S0716-078X2004000300010
20. Rice B, Westoby M. Evidence against the hypothesis that ant-dispersed seeds reach nutrient-enriched microsites. Ecology. 1986;67(5):1270–1274. https://doi.org/10.2307/1938682
21. Bond AWJ, Stock WD. The costs of leaving home: Ants disperse myrmecochorous seeds to low nutrient sites. Oecologia. 1989;81(3):412–417. https://doi.org/10.1007/BF00377092
22. Hirsch B, Kays R, Pereira VE, Jansen PA. Directed seed dispersal towards areas with low conspecific tree density by a scatter-hoarding rodent. Ecol Lett. 2012;15:1423–1429. https://doi.org/10.1111/ele.12000

Revisiting the peroneal trochlea of the StW 352 calcaneus

AUTHORS:
Ellison J. McNutt[1,2]
Alexander G. Claxton[3]
Kristian J. Carlson[4,5]

AFFILIATIONS:
[1]Department of Anthropology, Dartmouth College, Hanover, New Hampshire, USA
[2]Ecology, Evolution, Ecosystems and Society, Dartmouth College, Hanover, New Hampshire, USA
[3]Department of Anthropology, Boston University, Boston, Massachusetts, USA
[4]Department of Integrative Anatomical Sciences, University of Southern California, Los Angeles, California, USA
[5]Evolutionary Studies Institute, University of the Witwatersrand, Johannesburg, South Africa

CORRESPONDENCE TO:
Ellison McNutt

EMAIL:
Ellison.J.McNutt.GR@Dartmouth.edu

DATES:
Received: 06 Sep. 2017
Revised: 20 Oct. 2017
Accepted: 26 Oct. 2017

KEYWORDS:
Australopithecus africanus; Sterkfontein; peroneus longus; microCT; foot eversion

HOW TO CITE:
McNutt EJ, Claxton AG, Carlson KJ. Revisiting the peroneal trochlea of the StW 352 calcaneus. S Afr J Sci. 2017;113(11/12), Art. #2017-0306, 5 pages. http://dx.doi.org/10.17159/sajs.2017/20170306

ARTICLE INCLUDES:
× Supplementary material
× Data set

FUNDING:
Dartmouth College; National Research Foundation (South Africa); Department of Science and Technology (South Africa)

StW 352, from Sterkfontein Member 4 (South Africa), is a partial calcaneus attributed to *Australopithecus africanus* and is dated to ~2.0–2.6 Ma. The unusual robusticity of the peroneal trochlea (PT) of StW 352 has been commented on by several authors. The size of hominin PTs has been hypothesised to be positively correlated with the degree of recruitment of peroneus longus during bipedal locomotion and/or climbing. Given the potential functional relevance of an enlarged PT for reconstructing hominin activity patterns, we present the following previously unrecognised structural details of the reconstructed StW 352 that affect current interpretations of its functional morphology: (1) we estimate that the PT has been reattached to the body of the calcaneus ~5 mm dorso-distally from its original anatomical position; and (2) the presence of intrusive matrix has artificially misshaped the PT by expanding it laterally and proximodistally. Future studies of this specimen that apply geometric morphometrics, or other shape analysis tools, should compensate for these inaccuracies before undertaking comparisons between it and other calcanei. Additionally, given that the PT is likely smaller than previously reported for StW 352, caution should be exercised when using it to infer muscle function and extrapolate activity patterns of this individual, and thus by extension, within *Australopithecus africanus* in general. Lastly, these findings highlight the importance of not only the production of accurate reconstructions, but also the critical evaluation of the accuracy of existing reconstructions when working with damaged fossil material.

Significance

- This work epitomises the value of critically evaluating original fossil reconstructions, especially of postcranial elements.
- New technologies (e.g. microCT) offer non-destructive opportunities for evaluating/improving the accuracy of fossil reconstructions.
- Re-assessing StW 352 suggests peroneal muscles may have factored less prominently in *A. africanus* locomotion than previously thought.

StW 352 is a partial right calcaneus from Member 4 of the Sterkfontein Cave formation, South Africa[1,2] dated to approximately 2.0–2.6 Ma.[3] Because of its provenience in Member 4, StW 352 has typically been attributed to *Australopithecus africanus*[4-6], but that assignation is potentially complicated by a lack of consensus on the number of australopith species present in Sterkfontein Member 4[7-9].

The distal end of StW 352 is generally well preserved with the exception of damage to the plantar half of the articulation with the cuboid and a missing posterior tubercle. The latter is completely sheared off proximal to the posterior talar facet. There are two additional major breaks running through the calcaneal body, but both have been repaired with an unidentified type of glue. The first of these runs in a parasagittal plane and completely separates the peroneal trochlea from the rest of the calcaneal body. The second is approximately coronal in orientation, passing proximal to the sustentaculum tali, but distal to the posterior talar facet (see Figure 1). Given that the second break does not bisect the peroneal trochlea, it is likely that the trochlea was sheared off prior to the second break.

One prominent feature of the *A. africanus* StW 352 calcaneus is the unusual robusticity of its peroneal trochlea, as has been noted by several authors.[2,5,10,11] This robusticity is noteworthy because apes tend to exhibit larger peroneal trochleae than modern humans[12], and the peroneal trochlea of an *A. afarensis* calcaneus also has been described as 'massive'[13]. The peroneal trochlea is responsible for redirecting the tendons of peroneus longus and peroneus brevis travelling distally from the leg to their insertion in the foot.[14] The peroneal trochlea acts as the anchor point for the retinacula surrounding the tendons of these peroneal muscles and, thus, its size has been hypothesised to vary with their size and activity.[15]

Stern and Susman[16] used telemetered electromyographic studies of peroneus longus and peroneus brevis activation in *Pan troglodytes* and *Homo sapiens* to suggest that a large peroneal trochlea could have implications for understanding their functional role during bipedal locomotion and/or climbing in *Australopithecus*. They noted that both peroneal muscles were active during the second half of stance phase in humans, likely used to maintain the forefoot in relative eversion and assist in 'locking' the metatarsal joints while transferring weight medially towards the big toe.[16] In *P. troglodytes*, by contrast, neither peroneal muscle was active during terrestrial quadrupedalism and both seemed to be only variably active during stance phase of bipedalism.[16] More recent electromyographic studies of these peroneal muscles have observed activity in chimpanzees similar to that in humans during the last half of stance phase of bipedal gaits.[17] By comparison, both peroneal muscles were very active in chimpanzees during stance phase of quadrupedalism on vertical and horizontal simulated arboreal substrates, likely to aid in everting the foot on such supports.[16] Stern and Susman[16] hypothesised that the large (ape-like) peroneal trochlea in *A. afarensis* corresponded to large peroneal muscles reflective of a foot being used in an arboreal setting.[16]

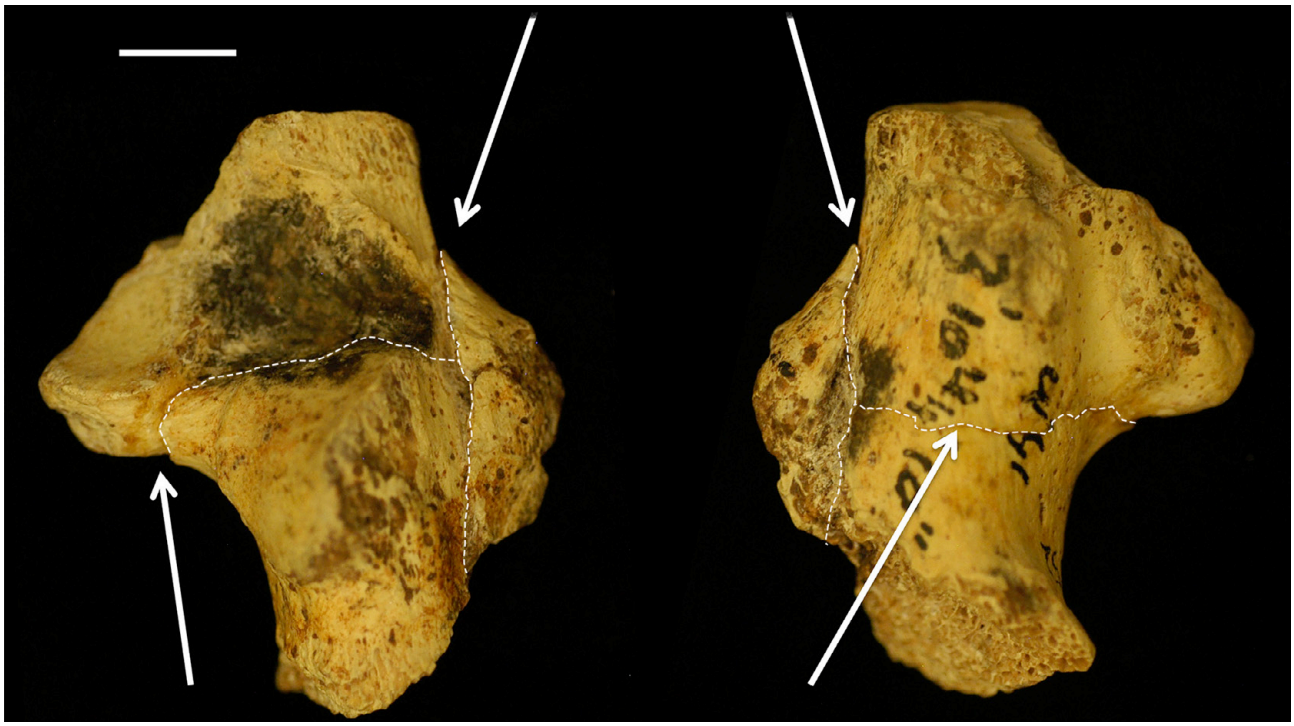


Figure 1: (Left) Dorsal and (right) plantar views of StW 352. Arrows and dashed lines highlight repaired breaks through the body of the calcaneus and separating the peroneal trochlea from the rest of the calcaneus. Note the lack of continuation of the 'horizontal' break through its intersection with the 'vertical' break in the illustration.

There are disagreements, however, about the degree to which arboreality was practised by *Australopithecus*.[13,18-21] If the size of the peroneal trochlea is correlated to frequent forceful contraction of the peroneal muscles, as has been suggested,[5,10,16] its position, size, and form informs in this debate. Given the uniqueness of the enlarged peroneal trochlea on StW 352, and thus its potential relevance for hominin functional morphology, we chose to carefully investigate its structural integrity using high-resolution computed tomography (CT).

Here, we critically evaluate StW 352's structural integrity, noting two issues that would affect current interpretations of the functional morphology of the reconstructed fossil. First, careful visual examination of the external surface of the peroneal trochlea suggests its proximodistal position may be incorrect. Upon close inspection, it appears that the peroneal trochlea has been reattached to the body of the calcaneus in such a way that there is cortical bone overlapping at both the dorsal and distal sides, leaving an artificial void inferior to the peroneal trochlea (see A and B in Figure 2). This malalignment suggests that the true attachment site should be more plantarly and proximally positioned compared to its current position. Second, the central region of the peroneal trochlea shows evidence of having been split and outwardly expanded by matrix infill, which appears to have spuriously augmented the overall robusticity of this anatomy, and its proximodistal dimension in particular (see C and D in Figure 2).

It is not uncommon to find evidence of taphonomic changes to the remains of fossil hominins from Sterkfontein Member 4. Multiple cranial and postcranial specimens from Sterkfontein Member 4 exhibit varying degrees of damage.[22-24] For example, many specimens show evidence of carnivore modification (i.e. bite marks), highlighting the important role of carnivores as accumulation agents in these deposits.[22] Thus, our suggestion of taphonomic or post-depositional modifications of StW 352 is not unusual for a specimen from Sterkfontein Member 4.

Visual examinations of external surfaces of the fossil were evaluated further using high-resolution CT. Because the objective here was to qualitatively examine structural integrity, specifically whether the trabecular struts and cortical shell within the fossil were correctly realigned, we compared similar high-resolution CT images from two representative extant hominoid calcanei. The comparative sample included an individual adult right calcaneus from a modern human (*H. sapiens*) and a chimpanzee (*P. troglodytes*), loaned from the Anthropology Department at Boston University and the Harvard Museum of Comparative Zoology, respectively. A chimpanzee calcaneus was included because chimpanzees represent the closest-living relatives to modern humans and have a peroneal trochlea that forms from a separate ossification site, as may have occurred in hominins with a large peroneal trochlea.[15] Chimpanzees also exhibit a generally more robust postcranial skeleton than modern humans, who appear to exhibit systemically lower bone mass, particularly with age.[25] Thus, by including the chimpanzee calcaneus, any gross differences in cortical and trabecular bone distribution, particularly those related to potential internal partitioning of the peroneal trochlea from the calcaneal body, generated by differences in the development of this anatomy should be accounted for in the current evaluation.

Image data from StW 352 were acquired using the industrial high-resolution CT scanner at the Evolutionary Studies Institute of the University of the Witwatersrand (Johannesburg, South Africa). Image data from the two comparative calcanei were acquired using the industrial high-resolution CT scanner at Harvard University's Center for Nanoscale Systems (Cambridge, MA, USA). Scan parameters for all specimens are listed in Table 1. Renderings produced from each image data set were sectioned in two planes through the peroneal trochlea in order to visualise the internal organisation: in a coronal plane (A) near its distal edge and in a transverse plane (B) through the dorsal surface (Figure 3). StW 352 was sectioned in a third plane: transversely (C) through the middle of the peroneal trochlea (Figure 4). There is an unusual partition of cortical bone separating the internal cavity of the peroneal trochlea of StW 352 from the internal cavity of the calcaneal body that is absent in the comparative material (A and B in Figure 3), both of which instead exhibit a single continuous internal cavity filled with trabecular struts extending from the lateral region of the calcaneal body into the peroneal trochlea. Section C (Figure 4) highlights the presence of intrusive matrix causing artificial outward (lateral) and proximodistal expansion of StW 352's peroneal trochlea.

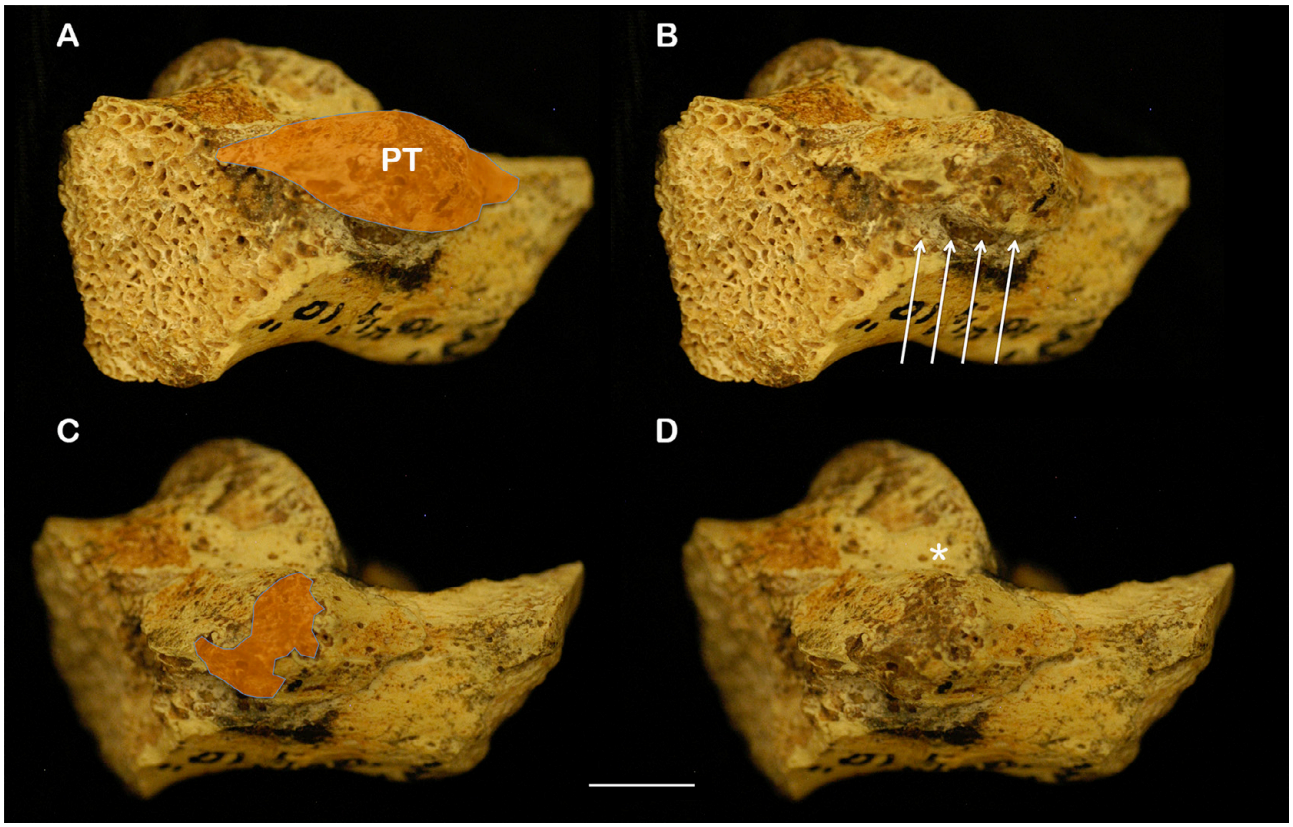


Figure 2: (A and B) Proximolateral views of StW 352. (A) Image highlighting the location of the peroneal trochlea (PT). (B) Arrows indicate the presence of a gap between the PT and the body of the calcaneus, which suggests that the true position of the PT should be more proximal and plantar than the current reconstructed position. (C and D) Lateral views of StW 352. (C) Highlighted outline of matrix infill. Notice how the proximal edge (to the left) may align fairly well with the dorsal edge (to the right), suggesting that these edges may refit if the intervening matrix was removed. (D) Unobstructed view of the splitting and expansion (underneath the star) of the middle of the PT caused by matrix.

Table 1: MicroCT scanning parameters for StW 352 and the comparative calcanei

Specimen number	Species	Scanning parameters				
		kV	μA	Voxel size (in μm)	Number of projections	Frames/second
BU 76	*Homo sapiens*	130	100	47.12	2958	1
MCZ 15312	*Pan troglodytes*	100	80	38.066	3092	1
StW 352[†]	*Australopithecus africanus*	90	135	29.6	5600	2

[†]Parameters taken from Zeininger et al.[6]

The high-resolution image data verify the initial visual examination of the external surface indicating that the peroneal trochlea in StW 352 has not been accurately reconstructed. Rather, it has been reattached to the main body of the calcaneus in an anatomically inaccurate position, and its robusticity (size and form) is at least partly enhanced by the remaining intrusive matrix. Its current form and the existing breakage pattern suggest that the peroneal trochlea must have been sheared away and proximodistally expanded before the coronally oriented break through the mid-body occurred. Some undefined period of time following this, the posterior tubercle appears to have been broken off and never recovered. Upon recovery of the existing fossil, the peroneal trochlea was reattached to the body approximately 5 mm dorso-distally (estimated from microCT image data) from its original anatomical position.

These observations have several implications. First, any future study of StW 352 employing geometric morphometrics, or other shape analysis tools, should compensate for the external displacement of its peroneal trochlea and use caution when commenting on the robusticity of its peroneal trochlea in comparative contexts. Moreover, a detailed digital reconstruction of the specimen would be advisable in order to more accurately reposition the peroneal trochlea and remove the excess matrix. Second, given the seemingly spurious outward expansion and proximodistal displacement observed, the actual size of the peroneal trochlea of StW 352 is likely smaller to an unknown extent and its anatomical position likely more proximal than previously described.[2,5,10,11] While correction of its size would lessen the robusticity of the peroneal trochlea to an uncertain precise extent, it would still appear to us to be relatively larger than typical human morphology. Nonetheless, because the size of hominin peroneal trochleae has been hypothesised to correspond with the degree of recruitment of the peroneal muscles,[5,16,18,26,27] and because this morphology is rarely preserved in the hominin fossil record, re-evaluation of StW 352 is crucial. Ultimately, we suggest that caution is warranted when using the present form of this bony anatomy on StW 352 to support interpretations of either an increased reliance on arboreal substrates or the peroneal muscles functioning as a possible compensatory mechanism for a less-developed medial arch during striding bipedalism in either this individual or, by extension, within *A. africanus* in general.

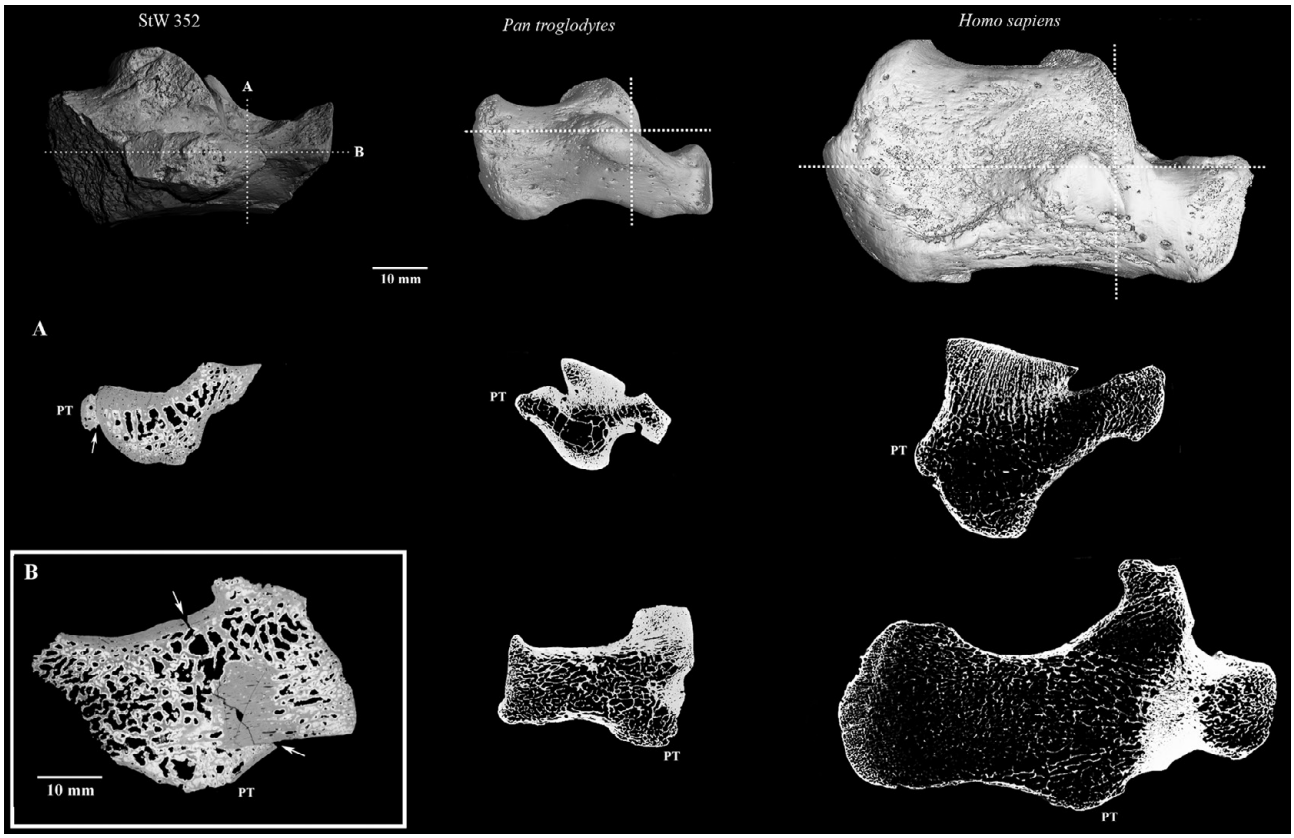


Figure 3: [From left to right] Top row: Lateral views of calcanei renderings from StW 352, *Pan troglodytes* and *Homo sapiens* showing the microCT sectional planes. Middle row: (A) Coronal section through the distal end of the peroneal trochlea (PT). Note the arrow indicating the presence of cortical bone under the PT of StW 352, but not in analogous locations in the comparative specimens. Bottom row: (B) Transverse section through the dorsal surface of the PT. Arrows indicate the location of both breaks through the body of StW 352 and the lack of similar cortical bone overlap in the comparative specimens.

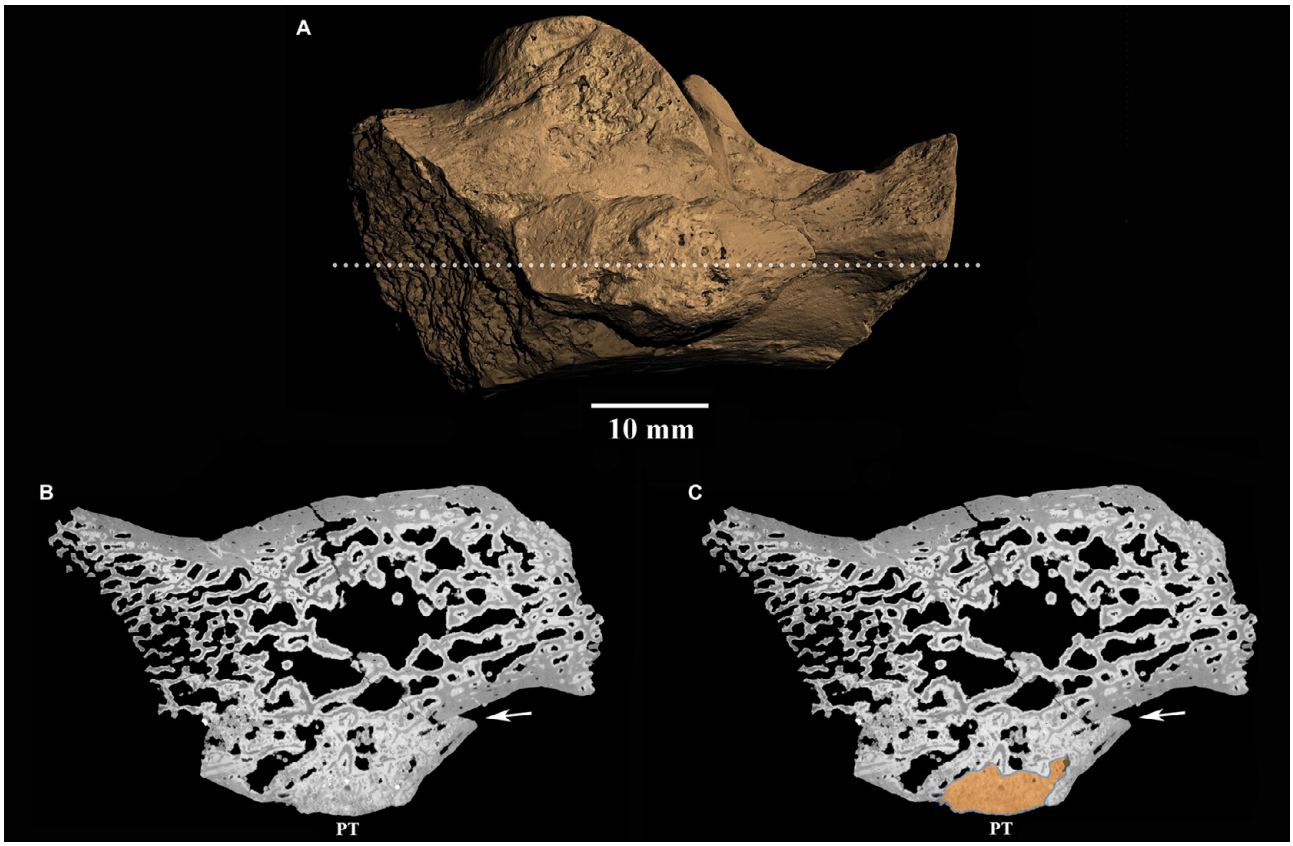


Figure 4: (A) Lateral view of StW 352 with the dotted line indicating the location of the transverse section through the middle of the peroneal trochlea (PT). (B) MicroCT image showing internal structure of the PT. Arrow indicates the area in which evidence of misalignment is best visualised. Note the presence of matrix infill located above the 'PT' label. (C) Repeat of image B with the shaded area demarking the approximate extent of the matrix infill.

Acknowledgements

We thank Dr Jeremy DeSilva for his advice and aid in funding E.J.M.'s travel to South Africa; the Department of Anthropology at Dartmouth College for funding travel for access to the original fossil material and the high-resolution scanning of comparative material in Boston; Dr Bernhard Zipfel and the Evolutionary Studies Institute of the University of the Witwatersrand for allowing access to StW 352 material; the National Research Foundation and the Department of Science and Technology (South Africa), for providing research funds to conduct high-resolution scanning of StW 352; Judy Chupasko and Mark Omura at the Harvard Museum of Comparative Zoology and the Department of Anthropology at Boston University for their assistance in loaning the comparative material. Scanning of comparative material was performed at the Center for Nanoscale Systems (CNS), a member of the National Nanotechnology Coordinated Infrastructure Network (NNCI), which is supported by the National Science Foundation under NSF award no. 1541959. CNS is part of Harvard University.

Authors' contributions

The study was conceived by E.J.M. The paper was written by E.J.M. with contributions and revisions by K.J.C. and A.G.C. Project design and data collection were completed by E.J.M., K.J.C. and A.G.C.

References

1. Deloison Y. Anatomie des os fossiles de pieds des hominidés d'Afrique du Sud datés entre 2, 4 et 3, 5 millions d'années. Interprétation quant à leur mode de locomotion. [Anatomy of the fossil bones of feet of hominids from South Africa dated between 2.4 and 3.5 million years. Interpretation as to their mode of locomotion]. Biométrie Hum Anthropol. Société de biométrie humaine. 2003;21(3–4):189–230. French.
2. McHenry HM. Early hominid postcrania: Phylogeny and function. In: Corruccini RS, Ciochon RL, editors. Integrative pathways to the past: Paleoanthropological papers in Honor of F. Clark Howell. Englewood Cliffs, NJ: Prentice-Hall; 1994. p. 251–268.
3. Pickering R, Kramers JD. Re-appraisal of the stratigraphy and determination of new U-Pb dates for the Sterkfontein hominin site, South Africa. J Hum Evol. 2010;59(1):70–86. https://doi.org/10.1016/j.jhevol.2010.03.014
4. DeSilva JM, Throckmorton ZJ. Lucy's flat feet: The relationship between the ankle and rearfoot arching in early hominins. PLoS One. 2010;5(12), e14432, 8 pages. https://doi.org/10.1371/journal.pone.0014432
5. Prang TC. Calcaneal robusticity in Plio-Pleistocene hominins: Implications for locomotor diversity and phylogeny. J Hum Evol. 2015;80:135–146. https://doi.org/10.1016/j.jhevol.2014.09.001
6. Zeininger A, Patel BA, Zipfel B, Carlson KJ. Trabecular architecture in the StW 352 fossil hominin calcaneus. J Hum Evol. 2016;97:145–58. https://doi.org/10.1016/j.jhevol.2016.05.009
7. Ahern J. Underestimating intraspecific variation: The problem with excluding Sts 19 from *Australopithecus africanus*. Am J Phys Anthropol. 1998;105(4):461–480. https://doi.org/10.1002/(SICI)1096-8644(199804)105:4<461::AID-AJPA5>3.0.CO;2-R
8. Clarke R. *Australopithecus* from Sterkfontein Caves, South Africa. In: Reed KE, Fleagle JG, Leakey RE, editors. The paleobiology of *Australopithecus*. New York: Springer; 2013. p. 105–123. https://doi.org/10.1007/978-94-007-5919-0_7
9. Grine FE, Delanty MM, Wood BA. Variation in mandibular postcanine dental morphology and hominin species representation in Member 4, Sterkfontein, South Africa. In: Reed KE, Fleagle JG, Leakey RE, editors. The paleobiology of *Australopithecus*. New York: Springer; 2013. p. 125–146. https://doi.org/10.1007/978-94-007-5919-0_8
10. Deloison Y. A new hypothesis on the origin of hominoid locomotion. In: Meldrum J, Hilton CE, editors. From biped to strider – the emergence of modern human walking, running, and resource transport. New York: Springer; 2004. p. 35–47. https://doi.org/10.1007/978-1-4419-8965-9_3
11. DeSilva JM. Vertical climbing adaptations in the anthropoid ankle and midfoot: Implications for locomotion in Miocene catarrhines and Plio-Pleistocene hominins [dissertation]. Ann Arbor, MI: University of Michigan; 2008.
12. Weidenreich F. The external tubercle of the human tuber calcanei. Am J Phys Anthropol. 1940;26:473–487. https://doi.org/10.1002/ajpa.1330260143
13. Latimer BM, Lovejoy CO, Johanson DC, Coppens Y. Hominid tarsal, metatarsal, and phalangeal bones recovered from the Hadar Formation 1974-1977 Collections. Am J Phys Anthropol 1982;57:701–719. https://doi.org/10.1002/ajpa.1330570412
14. Palmanovich E, Laver L, Brin YS, Kotz E, Hetsroni I, Mann G, et al. Peroneus longus tear and its relation to the peroneal tubercle: A review of the literature. Muscles Ligaments Tendons J. 2011;1(4):153–160.
15. Latimer B, Lovejoy CO. The calcaneus of *Australopithecus afarensis* and its implications for the evolution of bipedality. Am J Phys Anthropol. 1989;78(3):369–386. https://doi.org/10.1002/ajpa.1330780306
16. Stern Jr JT, Susman RL. The locomotor anatomy of *Australopithecus afarensis*. Am J Phys Anthropol. 1983;60(3):279–317. https://doi.org/10.1002/ajpa.1330600302
17. Larson SG, Stern Jr JT. Chimpanzee hind limb muscle activity patterns during bipedal walking [document on the Internet]. Stony Brook, NY: Stony Brook Primate Locomotion Laboratory; 2017 [cited 2017 October 19]. Available from: http://primatelocomotion.org/chimpanzee-bipedalism-project/overview-of-analyzed-results/chimpanzee-hind-limb-muscle-activity-patterns-during-bipedal-walking/#peroneus-longus
18. Susman RL, Stern Jr JT, Jungers WL. Arboreality and bipedality in the Hadar hominids. Folia Primatol. 1984;43(2–3):113–156. https://doi.org/10.1159/000156176
19. McHenry HM, Berger LR. Body proportions in *Australopithecus afarensis* and *A. africanus* and the origin of the genus *Homo*. J Hum Evol. 1998;35(1):1–22. https://doi.org/10.1006/jhev.1997.0197
20. Ward CV. Interpreting the posture and locomotion of *Australopithecus afarensis*: Where do we stand? Am J Phys Anthropol. 2002;119(S35):185–215. https://doi.org/10.1002/ajpa.10185
21. Green DJ, Gordon AD, Richmond BG. Limb-size proportions in *Australopithecus afarensis* and *Australopithecus africanus*. J Hum Evol. 2007;52(2):187–200. https://doi.org/10.1016/j.jhevol.2006.09.001
22. Pickering TR, Clarke RJ, Moggi-Cecchi J. Role of carnivores in the accumulation of the Sterkfontein Member 4 hominid assemblage: A taphonomic reassessment of the complete hominid fossil sample (1936–1999). Am J Phys Anthropol. 2004;125(1):1–15. https://doi.org/10.1002/ajpa.10278
23. Lockwood CA, Tobias PV. Morphology and affinities of new hominin cranial remains from Member 4 of the Sterkfontein Formation, Gauteng Province, South Africa. J Hum Evol. 2002;42(4):389–450. https://doi.org/10.1006/jhev.2001.0532
24. Vrba ES. A new study of the scapula of *Australopithecus africanus* from Sterkfontein. Am J Phys Anthropol. 1979;51(1):117–129. https://doi.org/10.1002/ajpa.1330510114
25. Chirchir H, Kivell TL, Ruff CB, Hublin J-J, Carlson KJ, Zipfel B, et al. Recent origin of low trabecular bone density in modern humans. Proc Natl Acad Sci USA. 2015;112(2):366–371. https://doi.org/10.1073/pnas.1411696112
26. Zipfel B, DeSilva JM, Kidd RS, Carlson KJ, Churchill SE, Berger LR. The foot and ankle of *Australopithecus sediba*. Science. 2011;333(6048):1417–1420. https://doi.org/10.1126/science.1202703
27. Harcourt-Smith WEH, Throckmorton Z, Congdon KA, Zipfel B, Deane AS, Drapeau MSM, et al. The foot of *Homo naledi*. Nat Commun. 2015;6:8432. https://doi.org/10.1038/ncomms9432