

JANUARY/FEBRUARY 2016

eISSN: 1996-7489



Applying scientific  
thinking in the  
service of society

ASSAf turns 20

Safe spending rates for  
South African retirees

Mitigating  
human–elephant  
conflict: responses of  
African elephants to  
a bee threat

Indicative hazard  
profile for strong  
winds in South Africa

Radiological and  
genetic analysis of a  
Late Iron Age mummy

# SOUTH AFRICAN Journal of Science



volume 112  
number 1/2

# SOUTH AFRICAN Journal of Science

volume 112

number 1/2

## EDITOR-IN-CHIEF

John Butler-Adam  
Office of the Vice Principal:  
Research and Graduate Education,  
University of Pretoria

## MANAGING EDITOR

Linda Fick  
Academy of Science of South Africa

## COPY EDITOR and PROOFREADER

Margaret Chamberlain  
Academy of Science of South Africa

## ONLINE PUBLISHING ADMINISTRATOR

Nadine Wubbeling  
Academy of Science of South Africa

## ASSOCIATE EDITORS

Nicolaas Beukes  
Department of Geology, University  
of Johannesburg

Tania Douglas  
Division of Biomedical Engineering,  
University of Cape Town

Menán du Plessis  
Department of General Linguistics,  
Stellenbosch University

Kavilan Moodley  
School of Mathematics, Statistics  
and Computer Science, University of  
KwaZulu-Natal

Jolanda Roux  
Forestry and Agricultural  
Biotechnology Institute, University  
of Pretoria

Pieter Steyn  
Department of Chemistry  
and Polymer Science,  
Stellenbosch University

Francis Thackeray  
Evolutionary Studies Institute,  
University of the Witwatersrand

Brian van Wilgen  
Centre for Invasion Biology,  
Department of Botany and Zoology,  
Stellenbosch University

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

Merridy Wilson-Strydom  
Centre for Research on Higher  
Education and Development,  
University of the Free State

## EDITORIAL ADVISORY BOARD

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

### Leader

ASSAf turns 20: Young enough to be dynamic and old enough to be trusted with its mission

*Wieland Gevers* ..... 1

### News and Views

South Africa debuts world-class Science Forum

*Edward W. Lempinen* ..... 3

Brave old world: Can today's university truly be 'home' to tomorrow's minds?

*Laurence Wright* ..... 5

What's new from the PAST?

*Robert J. Blumenschine & Andrea Leenen* ..... 7

Out of Africa: From *Homo naledi* to 'Homo cyborg'

*Gillian P. Christie & Derek Yach* ..... 10

### Scientific Correspondence

Safe spending rates for South African retirees

*Eben Maré* ..... 11

### Book Review

Casting the nets wide in the quest for ways to 'Africanise' research and teaching

*Menán du Plessis* ..... 15

What separates them from us?

*Rob Slotow* ..... 17

Whose knowledge? The politics of scholarship in South(ern) Africa:

A critical review of *Inside African Anthropology*

*Shahid Vawda* ..... 19

Participatory democracy for our time

*Anthony Egan* ..... 23

### Commentary

ASSAf: Promoting scholarly activity through the SAJS

*Roseanne Diab* ..... 25

Breast cancer: When do you stop reading the literature?

*Brenda Wingfield* ..... 27

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

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

**Published by**  
the Academy of Science of South  
Africa ([www.assaf.org.za](http://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](mailto:admin@sunbloem.co.za)

**Correspondence and  
enquiries**  
[sajs@assaf.org.za](mailto:sajs@assaf.org.za)

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

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

**Submissions**  
Submissions should be made at [http://  
mc.manuscriptcentral.com/sajs](http://mc.manuscriptcentral.com/sajs)

#### Cover caption

ASSAf celebrates its  
20th anniversary in 2016. In  
recognition of this milestone,  
the *South African Journal of  
Science* will be publishing  
commemorative articles  
throughout the year.

## Review Article

- Urban farming as a possible source of trace metals in human diets  
*Joshua O. Olowoyo & Gladness N. Lion* ..... 30
- Neutrophil extracellular traps and their role in health and disease  
*Jan G. Nel, Annette J. Theron, Roger Pool, Chrisna Durandt, Gregory R.  
Tintinger & Ronald Anderson* ..... 36
- The botanical content in the South African curriculum: A barren desert or a  
thriving forest?  
*Amelia L. Abrie* ..... 45

## Research Article

- Responses of African elephants towards a bee threat: Its application in  
mitigating human–elephant conflict  
*Mduduzi Ndlovu, Emma Devereux, Melissa Chieffe, Kendra Asklof & Alicia Russo* ..... 52
- High-speed detection of emergent market clustering via an unsupervised parallel  
genetic algorithm  
*Dieter Hendricks, Tim Gebbie & Diane Wilcox* ..... 57
- Publishing patterns at the Cape Peninsula University of Technology  
*Elisha Chiware & Lara Skelly* ..... 66
- Physicochemical characteristics of Bambara groundnut dietary fibres extracted  
using wet milling  
*Yvonne Maphosa & Victoria. A. Jideani* ..... 72
- Indicative hazard profile for strong winds in South Africa  
*Andries C. Kruger, Dechlan L. Pillay & Mark van Staden* ..... 80
- Radiological and genetic analysis of a Late Iron Age mummy from the Tuli  
Block, Botswana  
*Frank J. Rühli, Maryna Steyn, Morongwa N. Mosothwane, Lena Öhrström,  
Molebogeng K. Bodiba & Abigail Bouwman* ..... 91
- Radioactive nuclides in phosphogypsum from the lowveld region of South Africa  
*Xolani Msila, Frans Labuschagne, Werner Barnard & David G. Billing* ..... 98
- Investigating atmospheric photochemistry in the Johannesburg-Pretoria  
megacity using a box model  
*Alexandra S.M. Lourens, Tim M. Butler, J. Paul Beukes, Pieter G. van Zyl,  
Gerhard D. Fourie & Mark G. Lawrence* ..... 103
- Evidence for aeolian origins of *heuweltjies* from buried gravel layers  
*Michael D. Cramer, Johanna von Holdt, Lesego Khomo, Jeremy J. Midgley* ..... 114
- Australopithecus robustus* societies – one-male or multimale?  
*Katarzyna A. Kaszycka* ..... 124
- Visitors' views of human origins after visiting the Cradle of Humankind World  
Heritage Site  
*Anthony Lelliott* ..... 132
- Spoilage potential of a novel group of bacteria isolated from dairy products  
*L. Ingrid Tsôeu, Piet J. Jooste, George Charimba & Celia J. Hugo* ..... 140

## ASSAf turns 20: Young enough to be dynamic and old enough to be trusted with its mission

Some national science academies boast of their long histories, and (to adapt in a more positive direction Churchill's malicious gibe about the modesty of his political opponent Attlee) they mostly have a lot to be boastful about. If longevity is to be the main criterion on which the merits of an academy are to be determined, however, the case for starting a new one would be weak. The fact that the 'academy idea' has by now taken root in a majority of UN member nations, and the number still is increasing, shows that an alternative interpretation is correct: like universities, science academies have strong survival prospects in societies because they are in principle, and often in practice, a demonstrable 'public good'.

Again, as in the case of universities, making sure that a science academy is a real national asset requires considerable effort; the benefits do not simply fall from the sky.

A new academy that adopts and steadfastly maintains a fresh and contemporary approach to its mission within the core framework of practice can readily become a star performer. The argument will be made here that South Africa's national science academy has achieved this status, after only 20 years, despite having had to contend with many difficulties in its operating environment since its inception in 1996.

The process to establish the national academy – the Academy of Science of South Africa (ASSAf) – took about 5 years and was aligned with the momentous events that led to the first democratic election in South Africa in 1994. The nine-member planning team began its memorandum with the following set of assertions:

*Scientific thought and activity enrich us profoundly; they empower us to shape our living environment; they are keys that can open the doors to a peaceful and prosperous future. In a free society, an academy of sciences can be at once a symbol, an inspiration and a source of reliable counsel. It should take a form which is appropriate for the time and the place, allowing for further development through flexibility in its constitution. It should be fearless in its principal mission to respond with effective advice and action to our collective needs, dangers, opportunities and challenges.*

A way was then forged for the new, inclusive academy to be formed, designed to serve all the country's people as captured in the slogan 'science for society'. Parliament passed the ASSAf Act in 2001, 'licensing' ASSAf to receive public funding and to carry out its mission as the sole national science academy, also representing the country internationally in this arena.

ASSAf was intended by its founders (and by its parliamentary sponsors) to retain the best of the global academy tradition, but to be of this time and this place. Thus the constitution adopted by the nascent academy reflected an important principle that allowed ASSAf to jettison many out-of-date notions that were still carried forward in the academy tradition by older academies. Amongst these was the idea of academy fellowship or membership being a kind of reward for past academic efforts, a club of 'haves' which looked down on 'have-nots'. Another discarded viewpoint was one which regarded the word 'science' as synonymous with 'natural science' or 'hard science', the preserve of people who regarded themselves as the only 'proper' scientists. Yet another rejected approach was to regard science academies as merely bringing under

one privileged roof a number of different disciplines ('the sciences'), all constituting separately communities in each of which the constituent brains have 'constructed' themselves irreversibly into a unique mode of thought.

The newcomer, ASSAf, after considerable debate, and at some risk to its evolving support base, opted for the standpoint that a national 'science academy' should basically be devoted to the promotion and utilisation of the open-ended and evidence-based *way of enquiry* that is common to all empirical disciplines (hence ASSAf would be an 'Academy of Science', not of 'Sciences'). This approach meant that the distinctive powers of many disciplines would be harnessed to common purpose, at the highest level, to address societal problems – the principal mission of the organisation. The principle also made it logical (although still internationally unique) that elective membership of ASSAf would be based on the double criteria of excellence in science (across the entire disciplinary spectrum) *and* success in applying such high-level scientific thought for the benefit of society; it was thought that such scholars would find it easier to cross boundaries and relate to one another in a mutually respectful manner in a volunteer system of joint intellectual service.

ASSAf has published the century-old *South African Journal of Science* since 2002, and launched its science magazine *Quest* a few years later; both are multidisciplinary, and in their different ways are key vehicles for promoting the same cohesive principle espoused by ASSAf in its 'science-for-society' mission.

A further aspiration, present but not explicitly articulated in the founding decisions of ASSAf, was to avoid the gerontocracy so characteristic of older academies. This was partially addressed by the above-mentioned principles but given substantive form by the creation of the ASSAf-affiliated South African Young Academy of Science (SAYAS) in the second decade of ASSAf's existence.

ASSAf was also determined from the start to break down the infamous 'Limpopo curtain' that had prevented South African scholars from interacting with their northern neighbours and the rest of the continent – ASSAf wished to be part of a cooperative regional academy system. The opportunity to embed itself in this way came with the African Science Academy Development Initiative (ASADI) sponsored by the US National Academies from 2005 onwards, providing contact points, joint conferences and projects, and a substantial increase in the number of partner African science academies. The focus of mentoring provided through the initiative by the US National Academies was on best practice in the generation of evidence-based advice, and this assistance underpinned the central role of this activity for the African academies, including ASSAf. Towards the end of its first 20-year period, ASSAf was requested to host two African-region branches of international academy-linked bodies – the Regional Office for Africa (ROA) of the International Council for Science (ICSU), as well as the Regional Office for Sub-Saharan Africa of The World Academy of Sciences (TWAS). The 'Limpopo curtain' is certainly no more...

The value of a fresh approach to ASSAf's agenda was shown in various ways during its difficult setting-up period. In the decade before there was a significant track record of authoritative, evidence-based advice generated in the national interest, ASSAf in its core constituency of well-established researchers and scholars appeared, in traditional 'academy terms', to be junior to the two rival but largely sectoral bodies of the 'old South Africa': the Royal Society of South Africa (RSSAf) and Die Suid-Afrikaanse Akademie vir Wetenskap en Kuns (SAAWK), both of which have continued to exist after ASSAf's formation. During the

**HOW TO CITE:** Gevers W. ASSAf turns 20: Young enough to be dynamic and old enough to be trusted with its mission. *S Afr J Sci.* 2016;112(1/2), Art. #a0133, 2 pages. <http://dx.doi.org/10.17159/sajs.2016/a0133>



second half-century of its existence, the RSSAf had been in decline, with limited resources and influence. SAAWK, by contrast, functioned during the same period as the *de facto* national science academy of the country, one of the pillars of the Afrikaner-dominated state, recognised and funded by Parliament as a statutory organisation, extensively supported financially by the private sector, but involved in national policy formulation almost entirely 'off the record' and unaccountably. ASSAf, as a differently conceptualised newcomer whose primary purpose was to serve the whole society through scientific thinking across disciplines and not within disciplines, has by now fully established itself in its demonstrable commitment to transparency (all its reports are in the public domain), best consensus arising from multiple perspectives, and high quality as assured by independent and multiple peer review.

It is striking that some of the most influential leaders of advanced country academies began to advocate a shift from the traditional inward-looking focus of such bodies more or less at the time that ASSAf was being established, moving towards the same foregrounding of societal service in the form of consensus advice generated by a full review of available evidence across the disciplinary spectrum. ASSAf was recognised as having pre-aligned itself with this shift, believing firmly that a national science academy in the modern era exists primarily to make possible the efficient and effective mobilisation of a nation's intellectual 'firepower' to address its most urgent problems. It was consequently not a surprise that ASSAf was elected to membership of the first InterAcademy Council (IAC) when this was formed in 2000 to drive the performance of international consensus studies on issues of global importance.

ASSAf's first major consensus report on research publishing in and from South Africa (commissioned by government) was released in 2006 after extensive process guidelines had been developed for ensuring that the panel-based reviews were independent, reflected a best consensus of multiple perspectives, and were of a high scholarly standard (including thorough peer review). This report led to the launching of an ASSAf-led and government-funded programme to improve and support scholarly publishing in multiple ways, including the setting up of the National Scholarly Editors' Forum, acceptance of a National Code of Best Practice in journal editing and peer review, and the launching of a fully indexed, open-access e-platform (SciELO SA) already presenting the full-text of 60 local scholarly journals which have passed scrutiny by discipline-grouped ASSAf peer-review panels. This activity has had, and is still having, a significant and broad impact on scholarly practice in South Africa.

The advisory function espoused by ASSAf is by now also performed in various ways other than full consensus reviews. Well-organised and highly participatory forum-style workshops on problem areas can provide an indication relatively quickly of a 'beginning consensus' on priorities and possible solutions. 'Informed high-level consensus' opinions on key government strategies can be generated in short order by well-constituted expert panels if required urgently. Concise position papers can be released on matters of public controversy or confusion. ASSAf can also act as a channel within the country for the dissemination of consensus reports and advisories emanating from regional or global academy groupings or agencies, such as those produced by the IAC already mentioned. In every case, the ASSAf Council is charged with the final approval-and-release decisions, based on process correctness and scholarly quality: the Council is publicly accountable for maintaining the good reputation of ASSAf but does not 'second-guess' the findings and recommendations of its appointed panels.

One year after the groundbreaking research publishing report, ASSAf on its own initiative produced a consensus report on the evidence base concerning possible nutritional influences on the pandemic diseases caused by chronic HIV and *Mycobacterium tuberculosis* infection; the findings helped to end a disastrous period of 'HIV denialism' in South Africa and to initiate a concerted national programme of science-based healthcare to mitigate the damage to society and the economy that these diseases were causing. This report was widely reported internationally and acclaimed as a clear indication that Africa's science academies were independently capable of playing a significant role in addressing key issues affecting their societies.

Only 9 years after these two reports and many other consensus reviews, forum proceedings, advisories and position papers later, an ASSAf consensus report on policy issues concerning gender orientation in Africa, prepared in partnership with the Ugandan Academy of Sciences, was praised as courageous and timely in a lead editorial and feature article in *Nature*. The 'teenage academy' was now deemed worthy of frontline international attention.

In the face of this good record, it is finally necessary to discuss two caveats, the chief causes of concern on the part of ASSAf leadership and its supporters. One is the issue of independence, important for a body that needs government funding to maintain and build its infrastructure, as well as funding on a contractual basis for commissioned reviews or other forms of advice. The principle of accountability makes it appropriate that the use of taxpayer money by an independent (although statutory) academy should require formal proposals, budgets, financial reports and audits, and be assessed against performance. This unavoidably opens up the possibility of top-down control within a system where most other public entities are as clear-cut government agencies subject to such direct control, even within the parameters of their respective statutes. It stands to the immense credit of the government department responsible for ASSAf's public funding, the Department of Science and Technology, and the government more broadly, that they have appreciated the fact that the only good national science academy is an independent one, and have acted accordingly.

The second worry is the still inadequate realisation on the part of researchers in the public sector (universities and research councils), as well as those in the private sector, of the difference between the processes of prospective research into matters of importance, which may or may not create evidence for policy, and the systematic, consensus-seeking review of already available evidence by research-experienced experts with multiple disciplinary perspectives, in ways that are directly designed to provide an evidential basis for policy. While there is no *prima facie* reason why the second mode of investigation cannot be done in a university or research council setting, it is simply much better and more cost effective when it is performed by an academy which can effectively mobilise any number of appropriately equipped volunteer scholars from any number of skills areas, see to a high level of quality assurance, and ensure transparency, all in a fully accountable manner.

Science academies of the kind that ASSAf aspires to be will be judged mainly on their track records in assisting society. A promising start has been made, and one can justifiably be optimistic about a second 20-year period of high-level achievement in this sphere.

*Wieland Gevers is Professor Emeritus at the University of Cape Town and a past President of ASSAf (1998–2004).*



## South Africa debuts world-class Science Forum

### AUTHOR:

Edward W. Lempinen<sup>1</sup>

### AFFILIATION:

<sup>1</sup>The World Academy of Sciences, Trieste, Italy

### CORRESPONDENCE TO:

Edward Lempinen

### EMAIL:

elempinen@twas.org

### POSTAL ADDRESS:

TWAS, Enrico Fermi Building  
#107, Via Beirut 6, 34151  
Trieste, Italy

### KEYWORDS:

science and technology; global challenges; policymakers; big data; open data

### HOW TO CITE:

Lempinen E. South Africa debuts world-class Science Forum. *S Afr J Sci.* 2016;112(1/2), Art. #a0143, 2 pages. <http://dx.doi.org/10.17159/sajs.2016/a0143>

In the culture of science, knowledge is built not only by research in the lab and in the field, but also by sharing and reviewing that research. Conferences, therefore, have great importance: scientists come together and share their latest work; they debate it, and they plan future research. The model has evolved over hundreds of years.

But it has been given an injection of new energy in the past 15 years. Events such as EuroScience Open Forum and Japan's Science Agora not only bring scientists together to discuss the frontiers of research, they also create a powerful new dynamic by including policymakers, business leaders, students and the public in a decidedly international context. And now Africa has an impressive meeting of its own.

Science Forum South Africa, held 8–9 December 2015 in Pretoria, was a showcase for Africa's vision and strength before an audience that included high-level researchers and policymakers from Africa and around the world. It offered discussions on topics ranging from space science and climate change to education and science diplomacy. And it hosted the launch of a major global accord, 'Open Data in a Big Data World', by four leading international science organisations: the International Council for Science (ICSU), The World Academy of Sciences (TWAS), the International Social Sciences Council (ISSC) and the InterAcademy Partnership (IAP).

The central message was clear: Africa, with South Africa in a leadership role, is a committed partner in applying science and technology to regional and global challenges.

'Our primary rationale for this conference resides in our conviction that science, technology and innovation can and must play a central role in achieving sustainable development', said Naledi Pandor, South Africa's Minister of Science and Technology. 'Africa cannot advance without investing in science', Pandor added in her opening address. 'At present, there are efforts to enhance the status of science and to increase investment in research development and innovation. Unfortunately, science is still at the margins of government attention, seen as less significant than water scarcity, food security and disease burdens. Yet all of these can be addressed through science.'

Nkosazana Dlamini-Zuma, Chairperson of the African Union Commission, underscored that message in her keynote address. 'Research in agriculture and related fields, such as geology and climate, is critical to Africa's health', Dlamini-Zuma said. 'We are about science', she asserted. 'We have to ensure that we improve our intra-African trade, global trade, governance, infrastructure and security, all of which require science.'

Another key to strong science in Africa: providing influential roles for women and young people. Dlamini-Zuma called young people 'our biggest assets'. And she added: 'None of this will go anywhere if women are not involved.... They are actually more than 50% of the population, so if you leave them out of your development agenda, how do you think we will compete with those who are using all their potential and talent?'

### An African forum with global vision

Science Forum South Africa (<http://www.sfsa.co.za>) drew its inspiration from Japan's Science Agora and the EuroScience Open Forum, and all three were influenced by the American Association for the Advancement of Science (AAAS), which held its first science meeting in 1848. AAAS, EuroScience, the Japan Science and Technology Agency (which organises the Agora), and the Hungarian Academy of Sciences (organiser of the World Science Forum) were among the Forum partners.

In all, the Forum convened more than 1500 participants from 40 African nations, plus representatives from Asia, Europe, Latin America, North America and the South Pacific. There were four plenary sessions, 32 short seminars and 18 lectures. The forum, organised by South Africa's Department of Science and Technology (DST), was held at the Council for Scientific and Industrial Research (CSIR) Convention Centre.

The Academy of Science of South Africa (ASSAf) and South Africa's National Research Foundation (NRF) provided key support for the Forum. The African Union, ICSU, TWAS, the European Commission and the United Nations Organization for Education, Science and Culture (UNESCO) were among the organisations that sent high-level representatives.

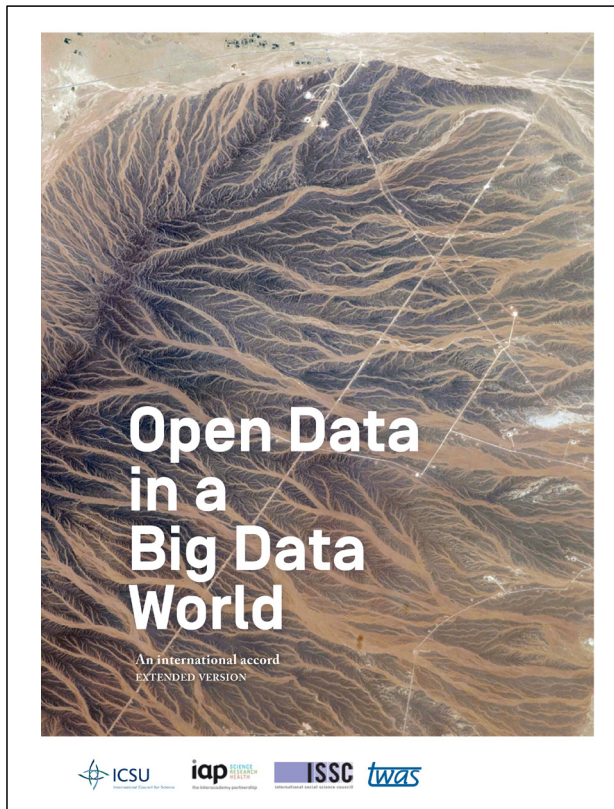
ICSU and TWAS each made significant contributions to the forum. ICSU President Gordon McBean and Executive Director Heide Hackmann each made comments at the opening ceremony; McBean later sat on a panel on climate change. TWAS Executive Director Romain Murenzi sat on panels focused on gender in science; education for science, technology, engineering and mathematics; and science advice to government. He then delivered closing remarks to thank the organisers, partners and participants.

### 'Open Data in a Big Data World'

With the open data accord, ICSU and TWAS expressed far-reaching ambitions, joined by ISSC and IAP. And it was clear at the Forum that they had strong allies at DST and among leading African scientists.

'As the data revolution accelerates and the scientific potential of big data becomes clearer, it is timely that the major representative bodies of international science promote the importance of open data,' said Geoffrey Boulton, President of CODATA, ICSU's Committee on Data, and leader of the working group that developed the accord. 'South Africa and other African nations have expressed support for the accord and the need to take full advantage of the transformative opportunities that are at hand.'

The accord, 'Open Data in a Big Data World', was the first project by the Science International partners. It now will be the focus of a year-long campaign to enlist endorsements and other support from science and policy bodies worldwide. It will also be the basis for a capacity-building initiative focused on Africa.



The open data accord can be retrieved at [www.icsu.org/science-international/accord](http://www.icsu.org/science-international/accord)

The four organisations have come together on open data at a historic moment: the digital revolution has created an explosion in the data available for analysis by scientists, policymakers and others. Extremely large data sets, or 'big data', are the engine of this revolution, and they can be used to discern and analyse powerful patterns in areas ranging from security and biodiversity to genetic research and human behaviour.

But the privatisation of knowledge poses a risk to the traditional conduct of research, and particularly to research in developing nations, they say. If big data sets that are the basis of research are not open and available, then other scientists will not be able to review the research to evaluate its conclusions.

'Open Data in a Big Data World' identifies the opportunities and challenges of the data revolution as an overarching interest for global science policy. It proposes 12 principles to guide the practice of open data, focused on the roles played by scientists, publishers, libraries and other stakeholders, and on technical requirements for open data. It also assesses the 'boundaries of openness'.

'Open data should be the default position for publicly funded science', according to the accord. 'Exceptions should be limited to issues of privacy, safety, security and to commercial use in the public interest. Proposed exceptions should be justified on a case-by-case basis and not as blanket exclusions.'

A further concern is that, without open data, developing nations in Africa and worldwide will be excluded from a vitally important new era of technology-driven research.

'Open access to data will be essential if developing countries are to join in the benefits of the big data revolution', said Murenzi. 'If developing nations are left behind, if they are unable to make a full contribution to the global research enterprise, that will be costly not only for them and their people, but for all nations.'

A key element of Science International is a big data/open data capacity-mobilising initiative for Africa. That project, already underway, is being led by ICSU and CODATA in concert with partners in Africa and other areas. It proposes the establishment of an African Open Data Platform, which will coordinate actions with national science systems.

The Science International partners represent more than 250 national and regional science academies, scientific unions and other organisations worldwide, with individual members at the highest levels of scientific research, policy and education. Science International is conceived as a series of annual meetings that bring together leaders of the four organisations and experts from around the world to address a key science policy challenge.

### 'Africa is the future'

The open data accord contributed to the substantive accomplishment of Science Forum South Africa. But the success of the Forum also was evident in its spirit.

The discussions and presentations included top African scientists and policymakers describing their work to address challenges in health, food security, education and other fields. Women played prominent roles throughout the Forum, and young South Africans were well represented on stage; in between sessions, they talked with scientists and policymakers in crowded halls and corridors. And the proceedings were imbued with the ideals of Nelson Mandela, the visionary political leader who died just 2 years earlier.

This combination of expertise, energy and idealism gave the Forum a powerful spirit of confidence about the present and optimism about the future. And this spirit was conveyed in extensive news coverage of the Forum sessions and the open data initiative.

In her opening address, Minister Pandor urged participants to take energy from the Forum and use it to build African science in the months and years ahead. They should urge government leaders to invest in science, and devote time to building innovation ecosystems. They must work to attract more pioneering science projects such as the Square Kilometre Array radiotelescope, which bring benefits across nations and regions. Global science collaboration should be encouraged. Young scientists should be supported. And science, she said, must reach out to society and encourage people 'to value the potential for development intrinsic to science, technology and innovation'.

In the closing session, science leaders reflected on the Forum's positive energy, with their focus on the future.

The Forum, Pandor said, was 'an extremely important occasion for our ambitions for science, technology and innovation'.

In remarks thanking Pandor and South Africa, Murenzi called the event 'historic'. 'For the first time', he said, 'an African nation has staged an international science forum – and it truly has attracted global attention. We have explored some of the most far-reaching issues of our time... All of these initiatives touch directly on sustainable development. All of these issues have a direct bearing on the health and prosperity of Africa, and of every nation on earth.'

These views were echoed by Peggy Oti-Boateng, a UNESCO Senior Programme Specialist for Science and Technology. The success of the Forum 'indicates that Africa has come of age', she said. 'Africa is the future.'





# Brave old world: Can today's university truly be 'home' to tomorrow's minds?

## AUTHOR:

Laurence Wright<sup>1</sup>

## AFFILIATION:

<sup>1</sup>Research Unit for Languages and Literature in the South African Context, North-West University, Potchefstroom, South Africa

## CORRESPONDENCE TO:

Laurence Wright

## EMAIL:

[l.wright@ru.ac.za](mailto:l.wright@ru.ac.za)

## POSTAL ADDRESS:

PO Box 2117, Sun Valley 7985, South Africa

## KEYWORDS:

higher education; reform; South Africa; creativity; knowledge producers

## HOW TO CITE:

Wright L. Brave old world: Can today's university truly be 'home' to tomorrow's minds? *S Afr J Sci.* 2016;112(1/2), Art. #a0136, 2 pages. <http://dx.doi.org/10.17159/sajs.2016/a0136>

Today, would you encourage your son or daughter to become an academic?

In the current international climate, I am not sure that I would. This confession is distressing, even heretical, coming from someone whose association with universities is lifelong and who is enjoying a very happy and reasonably productive academic career. What has gone wrong? I raise the issue because it matters. It matters not only for the future of South African universities but for universities everywhere.

It would be both simple and tedious to sketch the metamorphosis of universities internationally over the past 50 years. Key features are the increasing corporatisation, massification and managerialism of academic institutions; the routinisation and commodification of teaching and learning to accommodate growing numbers; the lowering of undergraduate standards in response to weak and struggling mass education systems; the suborning of the academic mission by disparate social factions seeking to influence society through other than intellectual means; the co-option of research by corporate, industrial and military funding agencies; and many similar developments. The outline is too well known to need repeating.

This is not a rant about returning to some mythical golden past. Universities around the world are as they are and have to cope, somehow, with the knapsack of challenges with which a mixed legacy of decision-making, good and bad, has landed them.

The real question universities face is whether they currently constitute the best – the most hospitable and effective – host institutions for those who wish to further human knowledge. Talented youngsters are voting with their feet. Can universities still confidently hope to attract and retain the best minds of succeeding generations, so that over time they advance the knowledge base of the societies which sustain them?

Frankly, if a significant percentage of an academic's time must perforce be devoted to teaching and assessing underprepared students who cannot even follow the material they ought to master; if the institution is weighed-down with bureaucracy; if decision-making processes are drawn-out and ineffective; if academic performance has to compete for scarce managerial attention with capital programmes, student protests and discipline, corporate fund-raising, community engagement, cultural transformation, institutional audits, performance reviews, and endless committee work, all hopelessly unrelated to that academic's research goals, why would anyone eager to shift the boundaries of human knowledge decide that universities are the appropriate place in which to do so? Looking around, the best and brightest may justifiably conclude that the university environment is no longer 'fit for purpose'.

As far as South Africa is concerned, I know of many high-calibre students (black and white) of recent decades, who have seriously considered an academic career, and then taken a cool, clear look at the prospects and decided against it. Some have headed for an academic career abroad (whether conditions there are better remains a moot point); others have decided that universities are no longer places conducive to the lives of intellectual innovation to which they aspire. This is a serious issue. Universities need talent – the talent in question no longer necessarily needs universities.

The world has changed. Today, multifarious social organisations compete for intellectual talent, some offering technical support and career prospects far superior to those available in an average university. Possibilities include migration from the university science park to state-of-the-art commercial and industrial development centres, which is a well-worn pathway for research scientists. Government-funded institutes offer attractive research prospects. Blue-sky think tanks absorb promising young researchers. Start-up companies are hungry for innovators. Moreover, it is not only the disciplines with obvious bausanic clout that proffer opportunities. A burgeoning array of national and international development agencies and NPOs offers appetising, hard-core research possibilities for humanities and social science graduates. Far from being 'unemployable', humanities majors have probably the greatest flexibility, earning power and variety of opportunity of any university graduate. Nothing compels them to choose a university career. Even those from that most marginal of disciplines (my own), literary studies, can build satisfying independent intellectual lives as writers, editors and media professionals outside the university. Writers of serious non-fiction earn very well, as do those producing school and university textbooks. For fine and performing artists, a university career remains only an uncomfortable fall-back position underwriting a financially insecure avocation whose ideal centre of gravity is elsewhere, in the public sphere – unless, that is, the university can win resources capable of sustaining a really first-rate environment for artistic creativity.

All these avenues point away from the university as the natural base for innovation and fresh knowledge-making in today's society. The alternative sites may carry their own risks and uncertainties, but they offer competitive rewards and 'lifestyle' attractions which few universities can match. None suffers the bureaucratic 'drag', confusion of aims, and institutional conflict characteristic of universities.

Not to be misunderstood: it is not politically motivated disruption I am complaining about. The increasing dysfunctionality of the university institution itself is what tends to work against effective research. It is often argued that the university talent drain is simply a matter of inadequate financial reward. Of course, academic remuneration cannot compete with government and industry. This is acutely the case in South Africa where a senior professor with several higher degrees, 15 years' work experience and much dedicated after-hours research may hope to earn about ZAR700 000 pa, the kind of salary freshly hatched graduates who score a deputy directorship in



government will earn from their first year. This is not wholly a bad thing. People whose main aim in life is to pile up personal wealth are not a natural fit with the academic enterprise.

But if universities cannot attract and retain budding research talent, their entire mission is at risk. There will always be warm bodies to fill academic posts, solid teachers who transfer knowledge and skills quite capably. As knowledge producers they are likely to deliver one research paper a year, often under protest. Real academic calibre is a very different matter. It is rare and unaccountable, and has to be nurtured and cherished in practical ways. These people are the game-changers. They thrive, not on being flattered and mollycoddled (the response of inept research managers who believe that prizes and financial reward are principally what encourages research), but on the opportunity of working in an intellectual environment wholly and intelligently focused on the knowledge enterprise. One cannot predict the arrival of such people, but one can provide for them when and if they appear.

What is to be done? One possible response, modelled on solutions adopted in the USA for an analogous predicament in the 1930s,<sup>1</sup> might be on the following lines:

1. Enhance graduate education by developing graduate schools, located in purpose-built buildings, offering taught modules by specialists shared nationally among graduate research centres.
2. Locate productive senior researchers primarily in these centres, with reduced teaching loads, to supervise graduates and pursue their own research. Staff could certainly include 'retired' researchers, but should consist primarily of high-achieving, mid-career researchers.
3. Build the scope of digital technology and technical support available to such researchers so that they can handle 'big data'.
4. Encourage the curatorial function of universities so that they house unique research material, and develop cohesive research investigations based on these data.
5. Cut through the confusion surrounding the roles of Deans and Research Managers. Deans are responsible for undergraduate education (to Honours level); Research Managers should mainly foster and support researchers and their students working in the graduate schools.<sup>2</sup>
6. Foster the kind of research environment where mentoring happens naturally through close collegiality, where research goals and ambitions are confidently articulated and shared, and where first-rate researcher-teachers are given the resources and explicit responsibility to develop the next generation of researchers.

Sounds wonderful, doesn't it? The trouble is, the change probably is not going to happen for some time, not in South Africa. Ideological exigencies aside, government subsidies are shrinking and will be used in the main to underpin undergraduate education. The economy is stagnating. Too many in South African society, quite understandably, view undergraduate degrees simply as a paper accreditation system, a lunch ticket. Student intakes are bulging with no signs of an increase in quality. There are no coherent moves to thoroughly educate would-be school teachers before they are trained as teachers.<sup>3,4</sup> It follows that little improvement in university environments can be anticipated in the short-to-medium term, which is unfortunate, because in that time we will lose many who might have become future intellectual stars.

Universities started because students wanted to learn from the leading intellects of their day. In this context we must ask whether they can still hope to attract and develop first-class teacher-researchers, or are they in the main heading for mediocrity? Are they really taking up the challenge of future research excellence? With brave exceptions, the answer is increasingly obvious. Wreathed in their time-honoured mystique, universities in general, and South African universities in particular, are beginning to lose the competition for young talent, and they need to re-examine their operational rationale and focus if this trend is to be reversed. Otherwise they will not attract their fair share of the best and brightest, and will consequently fail the country.

## References

1. Berelson B. Graduate education in the United States. New York: McGraw-Hill; 1960.
2. Seale O, Cross M. Why deans of universities need help to become better leaders. *The Conversation*. 2015 October 05; Education. Available from: <https://theconversation.com/why-deans-of-universities-need-help-to-become-better-leaders-48275>
3. Wright L. Education and the form of the humanities: An 'institutional remembering'. *Eng Acad Rev*. 2014;31(2):185–202. <http://dx.doi.org/10.1080/10131752.2014.965437>
4. Wright L, editor. South Africa's education crisis. Grahamstown: NISC; 2012.



## What's new from the PAST?

### AUTHORS:

Robert J. Blumenschine<sup>1</sup>  
Andrea Leenen<sup>1</sup>

### AFFILIATION:

<sup>1</sup>Palaeontological Scientific Trust,  
Johannesburg, South Africa

### CORRESPONDENCE TO:

Robert Blumenschine

### EMAIL:

rjb@pastafrica.co.za

### POSTAL ADDRESS:

PAST, PO Box 203,  
Parklands 2121, South Africa

### KEYWORDS:

Palaeontological Scientific  
Trust; All from One campaign;  
Walking Tall Educational Theatre  
Project; palaeoanthropology

### HOW TO CITE:

Blumenschine RJ, Leenen A.  
What's new from the PAST?  
S Afr J Sci. 2016;112(1/2), Art.  
#a0139, 3 pages. <http://dx.doi.org/10.17159/sajs.2016/a0139>

In 2003, Bob Brain<sup>1</sup> wrote in this journal about the first decade of operation of PAST, the Palaeontological Scientific Trust (then known as the Palaeo-Anthropology Scientific Trust). Entitled 'A perspective on the PAST', and written in his capacity as the NGO's scientific advisor, Brain described PAST's creation at the dawn of democracy in South Africa in 1994 as an effort to 'avert a serious financial crisis at the University of the Witwatersrand that was threatening the future of palaeoanthropological research there', particularly at the world-renowned Sterkfontein Caves. Brain outlined the central role PAST played in resuscitating student training and research in palaeoanthropology at Wits, such as supporting the discovery and development of the Little Foot skeleton at Sterkfontein,<sup>2</sup> through public sector financial support from Anglo-American, Standard Bank, and other corporations.

By 2003, Brain wrote that PAST was funding over 90% of human origins research in southern Africa. With continued corporate support, a growing and prominent board of trustees, and the infusion in 2002 of research funds from the South African Department of Science and Technology (DST) via the National Research Foundation (NRF), PAST expanded its support to the broader field of southern African palaeontology, including notably research in the richly fossiliferous Karoo Basin.

Importantly at this time, PAST's then fund manager, Andrea Leenen, established its now widely acclaimed Walking Tall Educational Theatre Project.<sup>1</sup> The project's origin sciences workshops were designed to meet the additional elements of PAST's founding mission: promoting public education into the sciences related to our origins, particularly among school children, and protecting South African hominid sites to facilitate palaeo-tourism.

In concluding his article, Brain characterised PAST's ambition as creating

*a proud and strong South African identity for each of its citizens as well as [bringing] its rich [fossil] heritage to the attention of the world...as part of South Africa's national drive towards unity and pride in our country and our continent.*

PAST has continued to grow in the 12 years since the publication of Brain's article. Its student training and research funding now extends into East and West Africa, and totalled ZAR4 million in grant awards in 2015. PAST is now the leading independent, African-based supporter of the sciences related to our origins on this continent.

As part of its public outreach mandate, PAST initiated in 2003 a keynote lecture series in collaboration with Standard Bank that has brought to South Africa world-renowned scientists such as Richard Dawkins, Tim White, Berhane Asfaw, Richard Leakey and Donald Johanson. Walking Tall has continued to grow, and has now reached over one million learners, educators and community members in southern Africa and, increasingly since 2011, in Tanzania, Kenya and Uganda. PAST also uses Walking Tall to illustrate human origins lectures that contribute to corporate and government diversity and transformation initiatives.

While retaining long-term support from Standard Bank, PAST's principal corporate sponsor, and the DST/NRF, PAST has expanded its donor base to now include JP Morgan Chase, the First Rand Foundation, the Oppenheimer Memorial Trust, the Claude Leon Foundation, South Africa's National Arts Council, and Investec, as well as a number of private individuals.

The Board of Trustees, now chaired by Rick Menell, continues to comprise prominent individuals from government and business as well as the scientific community. Andrea Leenen leads the PAST Executive as CEO, and was joined in 2011 by Robert Blumenschine as PAST's Chief Scientist.

Since that time, PAST has formalised its unique integration of research, student training, school-level education and public understanding by funding all of these activities at Human Heritage Hubs. These Hubs include the Sterkfontein Valley, Olduvai Gorge in Tanzania, the Turkana Basin in Kenya, and Ethiopia's Afar Rift. Funding of African-based scientists and their research remains at the core of PAST's mission, with the goal of helping to establish Africa as the global leader in the sciences related to our origins.

### The All from One campaign

In 2015, PAST greatly expanded its public understanding efforts with the launch of its global *All from One* campaign. Sponsored principally by Standard Bank, the DST, and the National Arts Council, the campaign uses the sciences underlying the shared origins of humankind and of all life forms to foster social cohesion and nature conservation.

The campaign recognises that the origin sciences can make a larger and more positive impact on society if they do not rely solely on the promotion of new fossil discoveries. While such discoveries capture the public's imagination, and provide the raw material of the origin sciences' service to society, individual discoveries have only a fleeting impact on public awareness. Further, media reports typically do not tie these discoveries into a broader view of the origin sciences, leaving the public with a series of disjointed findings that it is ill-equipped to integrate into meaningful social messages. Occasionally, if unrestrained by the lead discoverer, the media exaggerates the significance of fossil finds, breeding not only public scepticism of the discovery itself, but also cynicism toward the science in general. Such was the case with the South African media's coverage in 2015 of the hominid finds from the Rising Star Cave system in South Africa's 'Cradle of Humankind'.

*All from One* is based on the collective findings of a multitude of palaeontological and archaeological studies – and more recently ones in genetics – that document the origins and diversification of life and humankind. Because earlier findings about shared origins have stood the test of time, steadily gaining substantiation and meaning from

subsequent discoveries, the science underlying *All from One* does not require hyperbole to be compelling.

*All from One* can make a major contribution to society by using a basic understanding of shared origins to promote three messages. One of these messages – that Africa is the birthplace of humankind – echoes Brain's aspiration for PAST to contribute to unity and pride in South Africa and the continent overall.

Since Dart's<sup>3</sup> discovery of the Taung Child in 1924, a remarkable series of fossil and archaeological discoveries, coupled more recently with evidence from the comparative genetics of modern human populations, have demonstrated the African origins of all species directly ancestral to and including modern humans, starting with the emergence of the hominid lineage some 6 to 7 million years ago.

These findings bestow on Africa the status of being the primary storehouse of the world's fossil human heritage, starting with the earliest known hominid, *Sahelanthropus*. Even after the initial hominid colonisation of parts of Europe and Asia by early *Homo*, Africa appears to have remained the centre of human evolution, with the earliest known representatives of *Homo sapiens* emerging in Africa by some 200 000 years ago.<sup>4,5</sup>

Adding further to continental unity and pride is the abundance of evidence for the African origins of the fundamental characteristics that make us human and distinguish us from our closest living relatives, the chimpanzees. These findings, proceeding from the most primitive uniquely human traits to the most advanced, have steadily accumulated over the nine decades since Dart's recognition of bipedal adaptations in the Taung Child. Hence, discoveries at Olduvai Gorge of *Homo habilis* and the great antiquity of the Oldowan stone tool industry in the early 1960s<sup>6,7</sup> were the first to indicate that our genus, as well as the beginnings of humankind's dependence on technology, emerged in Africa. By the early 1980s, substantial evidence had accumulated to indicate that early *Homo erectus*<sup>8</sup> (a.k.a. *Homo ergaster*), and the more sophisticated stone technology of the earliest Acheulian<sup>9</sup>, also first appeared in Africa. At about the same time, discoveries in East Africa<sup>10</sup> and the South African cave of Swartkrans<sup>11</sup> suggested that the controlled use of fire was also an innovation that appeared in Africa. And, more recent discoveries, such as those from the southern Cape coast of Africa of the earliest known ornamentation and engravings<sup>12</sup>, now indicate that the advanced cognitive capabilities of *Homo sapiens* also arose on this continent.

*All from One* promotes a second socially relevant message emanating from our shared origins: the unity of humankind. This message is based not only on our long period of common ancestry outlined above, but also on the now abundant genetic<sup>13</sup>, palaeontological and archaeological<sup>14</sup> evidence for the relatively recent dispersal of *Homo sapiens* from Africa, beginning some 60 000 years ago. It is therefore within only the last 1% of the 6 million year history of the hominid lineage that population differences between Africans, Europeans, Asians and peoples from other regions of the world started to emerge – a finding reflected in the widely cited 99.9% similarity between the nucleotide sequences comprising the genomes of any two people.

*All from One* extends the concept of shared origins beyond humankind to include all life forms, extending back in time to the Last Universal Common Ancestor<sup>15</sup> that lived as many as 3.8 billion years ago. From this perspective, humankind represents one very small and recent branch among millions on the 'tree of life'<sup>16</sup> that is sustained in part by complex webs of ecological interdependence among species.

These ecological webs have been disrupted severely over the last 450 million years by a series of mass extinctions, during each of which at least 75% of known life forms went extinct in a relatively brief period of time. It is now increasingly clear that a sixth mass extinction is happening currently.<sup>17</sup> Extinction rates today are higher than in the last 66 million years, when the fifth mass extinction resulted in the demise of the non-avian dinosaurs and many other life forms. The ongoing wave of extinctions is widely and primarily attributed to human activities. If unabated, it could reach the 75% species-loss threshold

for a mass extinction in as few as 240 years<sup>18</sup>, leading to the loss of many of the essential ecosystems services<sup>17</sup> upon which humankind's survival depends.

The *All from One* campaign seeks to increase public awareness of our need to preserve earth's biodiversity. Public appreciation of the shared origins of all life forms should help to replace the common view that we exert dominion over nature with one that emphasises our dependence on it.

PAST is promoting the *All from One* campaign in several ways. The campaign's digital strategy centres on a campaign website, [www.past.org.za/AllfromOne](http://www.past.org.za/AllfromOne), through which individuals can commit to four human ideals arising from a consideration of shared origins: tolerance, unity, collaboration and conservation. Individuals can add their 'selfie' to a rapidly growing digital 'Face of Humanity' they can subsequently share on various social media platforms. A centrepiece of the campaign includes a like-named exhibition that describes the basic science underlying our shared origins. The physical exhibition opened in Johannesburg in November 2015 on the launch of the campaign, and will tour South Africa and eventually internationally, along with the Walking Tall Educational Theatre Project. A website version of the exhibition can be found at [www.past.org.za/Exhibition](http://www.past.org.za/Exhibition).

PAST invites people around the world to join *All from One* by using the basic science of our shared origins to promote tolerance of diversity so that we can together build an environmentally sustainable and prosperous future for all of humankind.

## References

1. Brain CK. A perspective on the PAST. *S Afr J Sci.* 2003;99:235–236.
2. Clarke RJ. First ever discovery of a well-preserved skull and associated skeleton of *Australopithecus*. *S Afr J Sci.* 1998;94:460–463.
3. Dart RA. *Australopithecus africanus*: The man-ape of South Africa. *Nature.* 1925;115:195–199. <http://dx.doi.org/10.1038/115195a0>
4. Cann RL, Stoneking M, Wilson AC. Mitochondrial DNA and human evolution. *Nature.* 1987;325:31–36. <http://dx.doi.org/10.1038/325031a0>
5. McDougall I, Brown FH, Fleagle JG. Stratigraphic placement and age of modern humans from Kibish, Ethiopia. *Nature.* 2005;433:733–736. <http://dx.doi.org/10.1038/nature03258>
6. Leakey LSB, Evernden JF, Curtis G. Age of Bed I, Olduvai Gorge, Tanganyika. *Nature.* 1961;191:478–479. <http://dx.doi.org/10.1038/191478a0>
7. Leakey LSB, Tobias PV, Napier JR. A new species of the genus *Homo* from Olduvai Gorge. *Nature.* 1964;202:7–9. <http://dx.doi.org/10.1038/202007a0>
8. Lepre CJ, Kent DV. New magnetostratigraphy for the Olduvai Subchron in the Koobi Fora Formation, northwest Kenya, with implications for early *Homo*. *Earth Planet Sci Lett.* 2010;290(3–4):362–374. <http://dx.doi.org/10.1016/j.epsl.2009.12.032>
9. Leakey MD. Olduvai Gorge. Volume 3. Cambridge: Cambridge University Press; 1971.
10. Clark JD, Harris JWK. Fire and its roles in early hominid lifeways. *Afr Archaeol Rev.* 1985;3:3–27. <http://dx.doi.org/10.1007/BF01117453>
11. Brain CK, Sillen A. Evidence from the Swartkrans Cave for the earliest use of fire. *Nature.* 1988;336:464–466. <http://dx.doi.org/10.1038/336464a0>
12. Henshilwood CS, d'Errico F, Yates R, Jacobs Z, Tribolo C, Duller GAT, et al. Emergence of modern human behavior: Middle Stone Age engravings from South Africa. *Science.* 2002;295:1278–1280. <http://dx.doi.org/10.1126/science.1067575>
13. Henn BM, Cavalli-Sforza LL, Feldman MW. The great human expansion. *Proc Natl Acad Sci USA.* 2012;109:17758–17764. <http://dx.doi.org/10.1073/pnas.1212380109>
14. Mellars P, Gon KC, Carr M, Soares PA, Richards MB. Genetic and archaeological perspectives on the initial modern human colonization of southern Asia. *Proc Natl Acad Sci USA.* 2013;110:10699–10704. <http://dx.doi.org/10.1073/pnas.1306043110>

15. Glansdorff N, Xu Y, Labedan B. The Last Universal Common Ancestor: Emergence, constitution and genetic legacy of an elusive forerunner. *Biol Direct.* 2008;3, Art. #29, 35 pages. <http://dx.doi.org/10.1186/1745-6150-3-29>
16. Sugden AM, Jasny BR, Culotta E, Pennisi E. Charting the evolutionary history of life. *Science.* 2003;300:1691. <http://dx.doi.org/10.1126/science.300.5626.1691>
17. Ceballos G, Ehrlich PR, Barnosky AD, Garcia A, Pringle RM, Palmer TM. Accelerated modern human-induced species losses: Entering the sixth mass extinction. *Sci Adv.* 2015;1, e1400253, 5 pages. <http://dx.doi.org/10.1126/sciadv.1400253>
18. Barnosky AD, Matzke N, Tomiya S, Wogan GOU, Swartz B, et al. Has the earth's sixth mass extinction already arrived? *Nature.* 2011;471:51–57. <http://dx.doi.org/10.1038/nature09678>





## Out of Africa: From *Homo naledi* to 'Homo cyborg'

### AUTHORS:

Gillian P. Christie<sup>1</sup>  
Derek Yach<sup>1</sup>

### AFFILIATION:

<sup>1</sup>The Vitality Institute,  
New York, NY, USA

### CORRESPONDENCE TO:

Gillian Christie

### EMAIL:

gchristie@thevitalitygroup.com

### POSTAL ADDRESS:

The Vitality Institute,  
3 Columbus Circle, New York,  
NY 10019, USA

### KEYWORDS:

palaeoanthropology; human  
origins; evolution; technology;  
cybernetics

### HOW TO CITE:

Christie GP, Yach D. Out of  
Africa: From *Homo naledi*  
to 'Homo cyborg'. S Afr J  
Sci. 2016;112(1/2), Art.  
#a0134, 1 page. [http://dx.doi.  
org/10.17159/sajs.2016/a0134](http://dx.doi.org/10.17159/sajs.2016/a0134)

No tools were unearthed, only traces of tool-using capabilities. No death record was chronicled, only sweeping estimations of several thousands or millions of years. No enhanced cognitive functioning was alleged, only insinuations of a primitive brain the size of an orange. As scientists hunted for fragments of our earliest life forms in underground cave systems, they may also have uncovered traces of our not-so-human futures.

The earth-shattering discovery of a new species of hominin by palaeontologists operating in South Africa's underground cave system enhances our understanding of our human origins. After six 'underground astronauts' treaded through chambers of dripping stalactites and cracks of limestone to uncover *Homo naledi*, the piecing of 1500 preserved bones from at least 15 individuals confirms once again that our early roots are in Africa.

Other early roots were further explored only a few months after the discovery of *Homo naledi*, with the stunning find of 47 human teeth deep inside a cave system in Daoxian, China. Dated between 80 000 to 120 000 years old, the discovery of the teeth leads us to question only one migration out of Africa – could there have been multiple waves of movement off the continent?

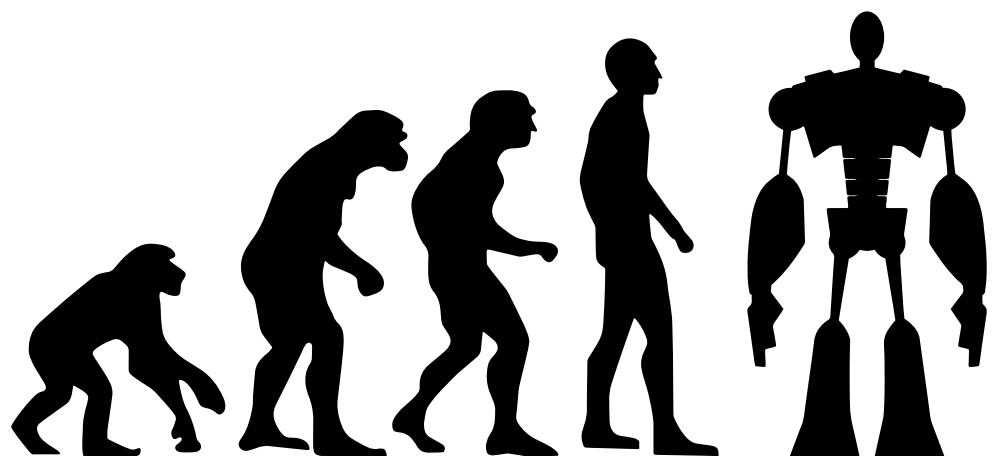
The uncovering of these ancient fossils undoubtedly represents a victory for palaeoanthropology. It also symbolises a technological triumph. In South Africa, petite cavers – all women – were recruited via social media, and their journeys were scrutinised through blogs and video diaries, complemented with Twitter and Facebook updates. Electrical cords dangled through caves, cameras above ground enabled real-time viewing below, and the recovery of fossils required 3D scanning equipment. Technology was also used to convene more than 50 experienced scientists and early-career researchers to categorise and report on the skeletal material. In China, sophisticated technologies were used to determine that the teeth contained no radioactive carbon, which led to the identification of their age using calcite deposits. Technological advancements thread through this tale.

Despite the technological prowess used for the discoveries, the certainty of our *Homo* origins disintegrates with time: Is *Homo naledi* 20 000 or 2 million years old? Are the teeth 80 000 or 120 000 years old? Were the bodies disposed of deliberately or caught in a death trap scenario? Scientists suggest it may be decades before convincing responses emerge for either scenario.

Within our lifetime, *Homo naledi* and its descendants – us – could plausibly evolve into 'Homo cyborg'. Techniques for human enhancement already exist to augment our mental and physical capabilities. Our bodies are restored to their original states through artificial medical remedies: pacemakers are routinely embedded in us to preserve, restore, or replace heart beating functions; and paralysed individuals move robotic limbs by channelling thoughts through electronic implants. The manifestation of *Homo cyborg* advances medicine to enhance and reshape the natural body and its human functions.

The convergence of computing and neuroscience will further propel the notion of *Homo cyborg*. The technology futurist Ray Kurzweil predicts that our neocortex will connect to the cloud by 2030 – a mere 15 years away. As our cognition fades with time, machines will assume lost functions. The human brain will die, but our minds will live on, perhaps even in a machine. We will be able to remedy biological demise through the life of our mind using technology. Humans risk losing control of their evolution. Artificial bodies could prevail.

Norbert Wiener, the US mathematician and founder of cybernetics, cautioned that technological advancement could result in the dehumanisation of humans. We employed cutting-edge technology to assist us in uncovering our ancestors. For our future tomorrow, we must balance our human touch with technological sophistication. We may not entirely understand our past out of Africa, but we certainly hold the tools to our future.



# Safe spending rates for South African retirees

**AUTHOR:**

Eben Maré<sup>1</sup>

*When a man retires and time is no longer a matter of urgent importance, his colleagues generally present him with a watch*

R.C. Sherriff

**AFFILIATION:**

<sup>1</sup>Department of Mathematics and Applied Mathematics, University of Pretoria, Pretoria, South Africa

Retirees face a difficult financial choice. They need to decide on an appropriate amount to withdraw from their retirement investment and they need to ‘optimise’ between outliving their money and living below their means. In the USA, financial advisors frequently advocate a ‘safe’ 4% withdrawal (or spending) rate. We examine spending rates for retirees in a South African setting.

**CORRESPONDENCE TO:**

Eben Maré

‘We want hope for riches and freedom from the fear of poverty’ wrote Statman<sup>1(p.107)</sup>. These words hold very true in retirement. Huxley and Burns<sup>2(p.203)</sup> note: ‘Three goals: Don’t go broke, preserve the Corpus, or preserve buying power.’

**EMAIL:**

Eben.Mare@up.ac.za

In an influential study, Cooley et al.<sup>3,4</sup> investigated the impact of various withdrawal (or spending) rates from a retirement portfolio while keeping these goals in mind. They note<sup>3(p.16)</sup>:

**POSTAL ADDRESS:**

Department of Mathematics and Applied Mathematics, University of Pretoria, Private Bag X20, Hatfield 0028, South Africa

*Most investors who plan for retirement eventually confront the question of how much money they should plan to withdraw from their investment portfolio. The dilemma is that if they withdraw too much, they prematurely exhaust the portfolio, but if they withdraw too little, they unnecessarily lower their standard of living.*

**KEYWORDS:**

retirement; asset allocation; withdrawal; spending; portfolio success rate

As a rule of thumb their conclusions are often used by advisers as the so-called ‘4% safe withdrawal rate’.

**HOW TO CITE:**

Maré E. Safe spending rates for South African retirees. S Afr J Sci. 2016;112(1/2), Art. #a0138, 4 pages. <http://dx.doi.org/10.17159/sajs.2016/a0138>

In South Africa, retirees typically choose between so-called single life guaranteed annuities and living annuities with their retirement proceeds. A single life guaranteed annuity is an insurance contract which covers the insured pensioner for life and yields a defined income. So-called living annuities allow the pensioner freedom to invest in a wide spectrum of investment vehicles while drawing a monthly amount for pension – currently limited by law to be between 2.5% and 17.5% per annum.

In guaranteed annuity products the insurer bears the longevity and investment risks. In contrast, living annuities allow retirees to benefit fully from their investment asset allocation. Furthermore, the balance of assets accrues to dependants upon death.

The 4% annual safe withdrawal rate studies by Cooley et al.<sup>3,4</sup> were performed for US retirees. The South African situation is different; typically as a result of different mortality rates, spending patterns as well as different asset class returns. South African inflation is structurally higher than inflation in the USA. Retirees in South Africa spend more aggressively than do those in the USA, typically drawing in excess of 6% of their portfolios annually as pension.

Here we examine withdrawal rates (i.e. spending rates) in living annuities for South African retirees.

Let’s assume one retires at age 60. Based on 2012 South African life tables obtained from the World Health Organization, the conditional survival probability table shown as Table 1 can be calculated.

**Table 1:** Conditional survival probability given age 60

Conditional survival probability given age 60 to age:	Men	Women	Both sexes
65	82%	90%	87%
70	61%	78%	70%
75	40%	62%	52%
80	22%	45%	35%
85	12%	27%	20%
90	5%	13%	10%
95	2%	5%	3%
100	0%	1%	1%
Conditional median life expectancy given age 60	73	78	76

South African median life expectancy hovers around 64 years (for both sexes) – the conditional median life expectancy, given age 60, is 76 years for both sexes. It is clear from Table 1 that if pensioners reach their retirement years they may be facing 15–20 more years of life with substantial probability; note that conditional life expectancy increases with age.

The retiree faces this question: How much can I spend on a periodic basis, given a specific retirement investment portfolio, to avoid running out of money?

To align our analysis with the work by Cooley et al.<sup>3,4</sup>, we consider an asset allocation between bonds and equities. In Table 2 we highlight the long-term investment returns of these investment classes on a nominal and real (i.e. inflation-adjusted) basis.

**Table 2:** Comparison of US and South African asset class returns

Asset class returns 1900–2014 (arithmetic mean returns % pa)			
	Asset class	USA	RSA
Nominal	Equities	11.5	14.8
	Bonds	5.4	7.3
	Inflation	3.0	5.2
Real	Equities	8.5	9.5
	Bonds	2.5	2.4

Source: Dimson et al.<sup>5</sup>

We consider an investment portfolio with a fixed asset allocation between bonds and equities and draw an income from the portfolio on a monthly basis (income is adjusted for inflation, i.e. we look at real spending rates). Cooley et al.<sup>3,4</sup> and Bengen<sup>6</sup> use a methodology of overlapping periods (also referred to as rolling periods) to calculate end-of-period portfolio values from equities and bond returns. Cooley et al.<sup>7</sup> also consider a Monte Carlo based simulation based on distributional characteristics of the asset classes.

Here the methodology is based on random sampling (with replacement) of monthly South African bond and equity total returns (i.e. we assume full reinvestment of interest and dividend proceeds) considered over the period 1950 to 2014 (data source: I-NET/Bridge). This methodology is referred to as a bootstrap simulation.

We consider the bootstrap simulation over fixed investment periods of 15, 20, 25 and 30 years, assuming no mortality. The monthly portfolio value is therefore a function of the simulated investment returns reduced by the amounts withdrawn. We perform 10 000 simulations over all fixed investment periods.

As in Cooley et al.<sup>3</sup>, we consider a portfolio successful if it has capital left at the end of the specific investment period considered and report the portfolio success rates (i.e. the percentage of successful portfolios based on the bootstrap simulations).

Table 3 details the results. As an example, to aid interpretation, the success rate for a portfolio consisting of 50% stocks and 50% bonds considered over an investment period of 30 years is 35% if we plan to draw 7% pension annually. Recall that our analysis adjusts for the effects of inflation hence the 7% is the real withdrawal rate.

**Table 3:** Portfolio success rate for different asset allocation and withdrawal rates

Payout Period	Withdrawal rate as percentage of initial investment value									
	3%	4%	5%	6%	7%	8%	9%	10%	11%	12%
<b>100% Stocks</b>										
15 Years	100%	99%	95%	93%	82%	78%	67%	60%	48%	39%
20 Years	99%	96%	90%	81%	73%	61%	49%	41%	38%	24%
25 Years	96%	92%	82%	76%	65%	55%	41%	32%	26%	19%
30 Years	96%	89%	78%	71%	56%	47%	37%	29%	23%	17%
<b>75% Stocks/25% Bonds</b>										
15 Years	100%	99%	98%	93%	88%	77%	64%	51%	39%	29%
20 Years	99%	98%	91%	84%	73%	58%	42%	28%	22%	14%
25 Years	98%	94%	85%	75%	57%	43%	28%	24%	14%	7%
30 Years	98%	91%	75%	61%	49%	35%	23%	14%	10%	7%
<b>50% Stocks/50% Bonds</b>										
15 Years	100%	100%	99%	97%	85%	69%	50%	31%	21%	9%
20 Years	100%	98%	94%	81%	61%	47%	26%	15%	7%	2%
25 Years	100%	95%	87%	65%	44%	24%	12%	5%	2%	2%
30 Years	99%	89%	73%	51%	35%	17%	7%	3%	2%	0%
<b>25% Stocks/75% Bonds</b>										
15 Years	100%	100%	100%	97%	81%	56%	32%	12%	4%	1%
20 Years	100%	99%	92%	71%	42%	16%	4%	2%	0%	0%
25 Years	100%	93%	71%	38%	16%	6%	1%	0%	0%	0%
30 Years	99%	84%	51%	24%	6%	2%	0%	0%	0%	0%
<b>100% Bonds</b>										
15 Years	100%	100%	99%	89%	61%	27%	10%	2%	1%	0%
20 Years	100%	96%	75%	37%	11%	2%	1%	0%	0%	0%
25 Years	98%	76%	34%	10%	2%	0%	0%	0%	0%	0%
30 Years	86%	46%	14%	2%	0%	0%	0%	0%	0%	0%

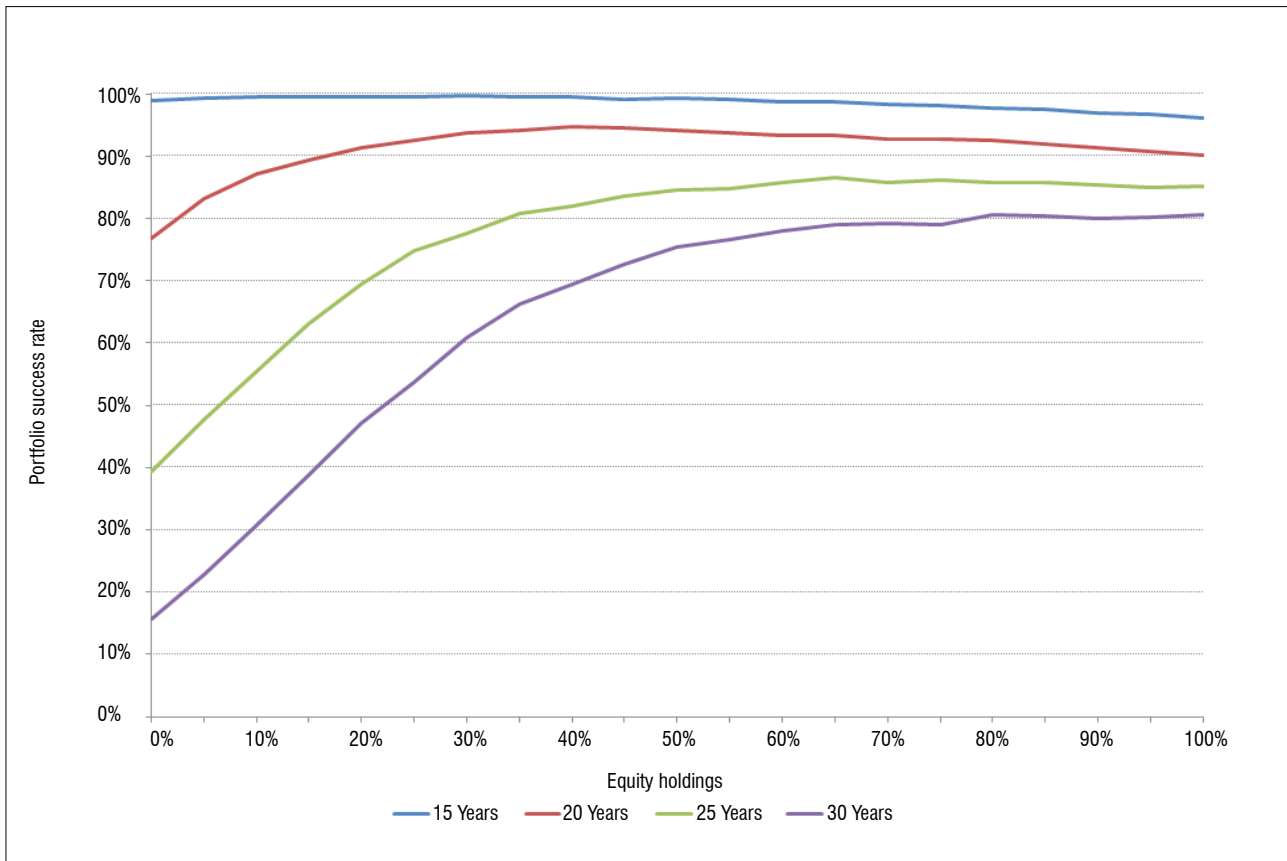


Figure 1: Portfolio success rate for 5% withdrawal rate.

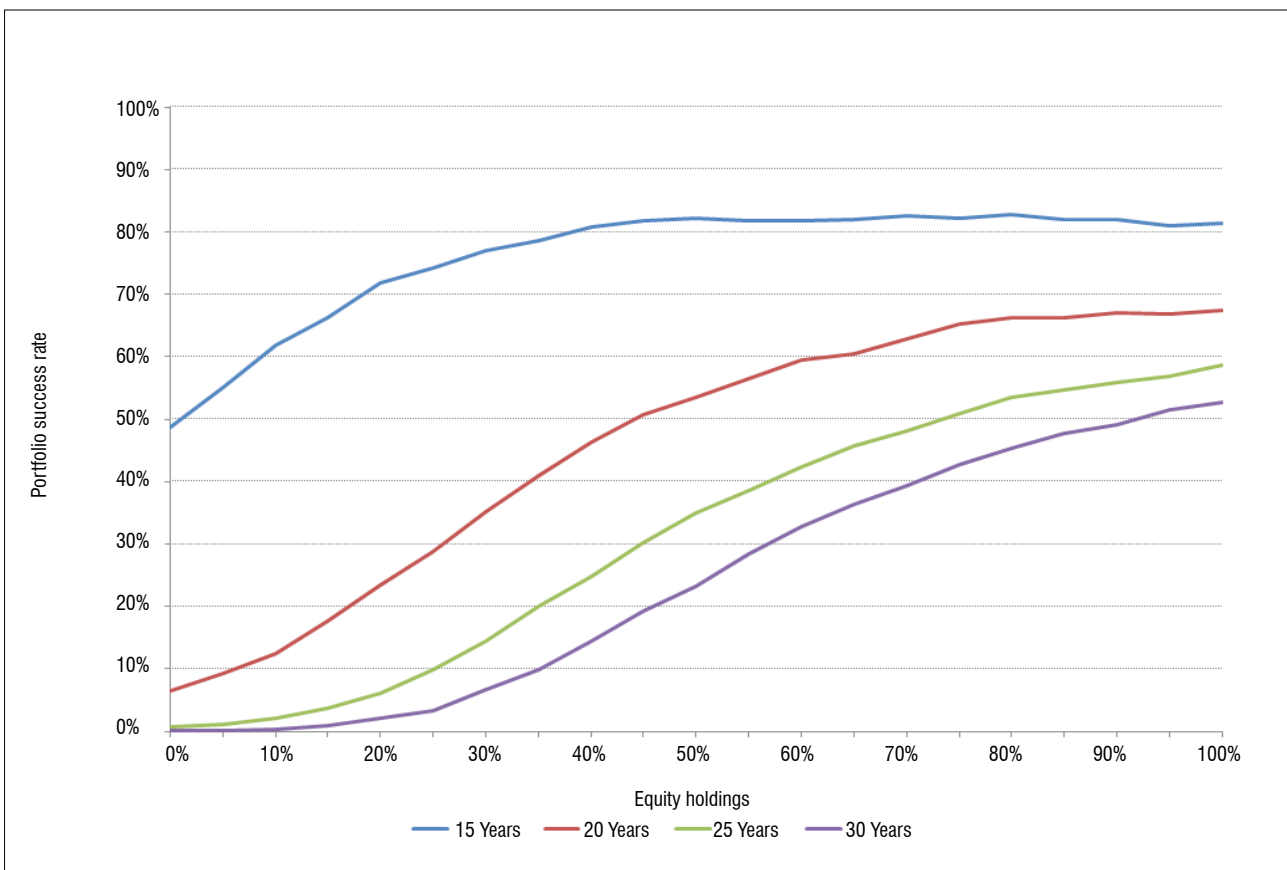


Figure 2: Portfolio success rate for 7.5% withdrawal rate.



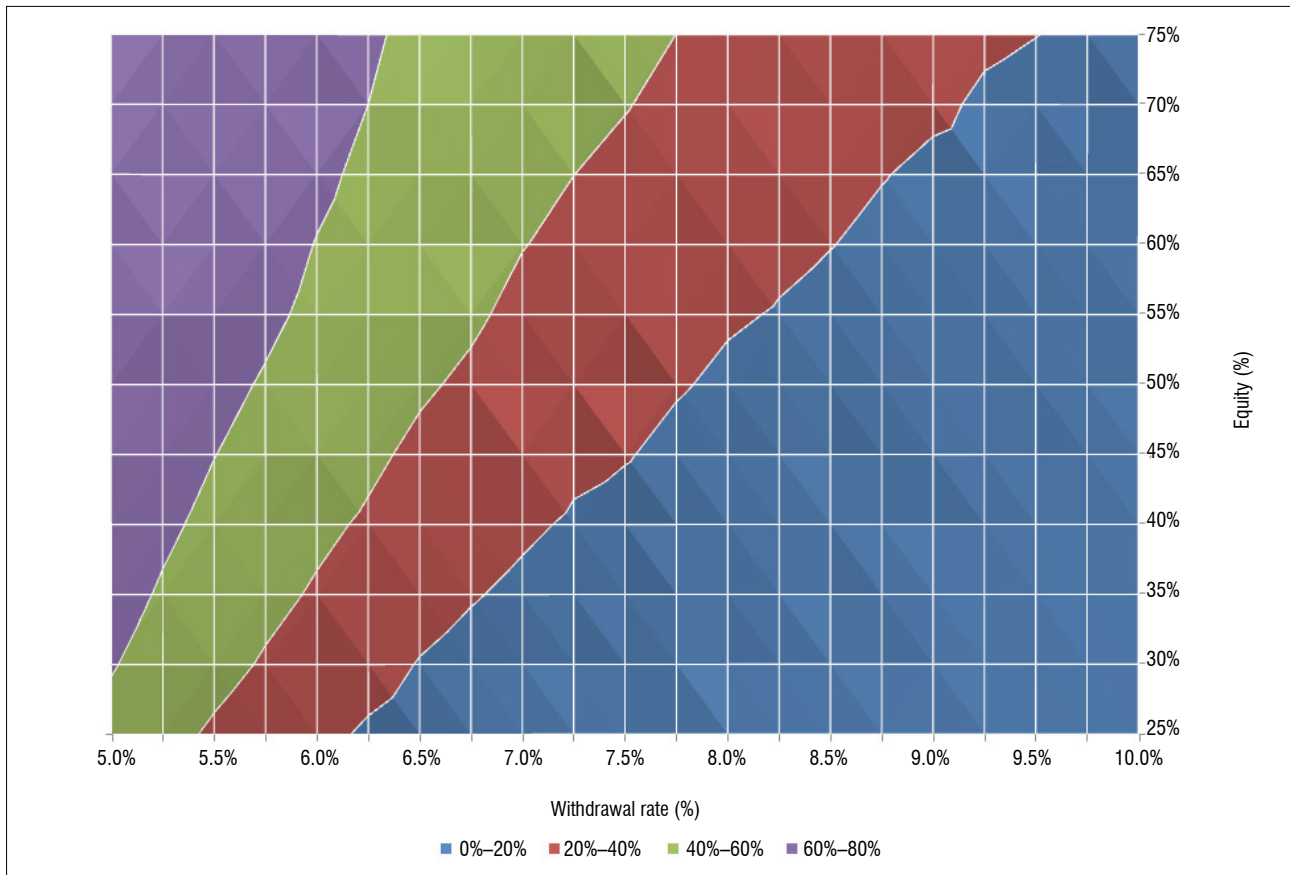


Figure 3: Portfolio success rate as a function of the withdrawal rate and equity holding.

We note the effects of high withdrawal rates on the portfolio success rates – even over relatively short investment periods (15 years). Over long investment periods (30 years), even low withdrawal rates might not protect us fully against running out of capital (Figures 1 and 2.)

The results detail the effects of asset allocation clearly. A 30-year portfolio consisting only of bonds will have a 2% success rate given a withdrawal rate of 6% annually, whereas the same analysis for a portfolio consisting only of equities yields a 71% success rate. Figure 3 details the 30-year portfolio success rate as a function of withdrawal rate versus equity holding. The results show very clearly, in the South African context, that spending rates above 6% yield low portfolio success rates (less than 60%) for moderate equity holdings. This conclusion can be drawn from Figures 1 and 2 as well.

In our approach we consider the distribution of terminal account balances and also calculate the probability of the investor’s capital being depleted. This corresponds to the notion of a safe withdrawal rate.

In the South African context it would appear that a 5% withdrawal rate is sustainable over relatively short periods (15 years or less). Longer-term investment periods yield results which are similar to the Cooley et al.<sup>3,4</sup> studies.

These results are important. Scott et al.<sup>8</sup> criticised the notion of withdrawing a fixed real amount from an inherently volatile portfolio. I concur; the results obtained here indicate moderation and a cautious approach – the risk associated with the investment environment is

significant – while one can expect the risky nature to be maintained in future years there can be no guarantee that historical returns would be maintained. I quote from Cooley et al.<sup>7(p.115)</sup> in conclusion: ‘A portfolio is only successful if it lasts as long as required by the retiree.’

## References

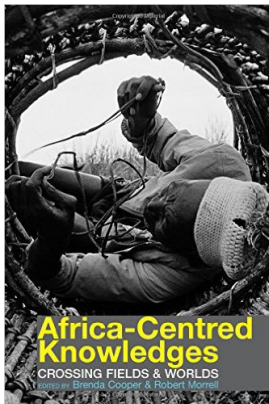
1. Statman M. What investors really want. New York: McGraw-Hill; 2011.
2. Huxley SJ, Burns JB. Asset dedication: How to grow wealthy with the next generation of asset allocation. New York: McGraw-Hill; 2005.
3. Cooley PL, Hubbard CM, Walz DT. Retirement savings: Choosing a withdrawal rate that is sustainable. *J Am Assoc Individual Investors*. 1998;20(2):16–21.
4. Cooley PL, Hubbard CM, Walz DT. Sustainable withdrawal rates from your retirement portfolio. *Financ Counsel Plan*. 1999;10(1):40–50.
5. Dimson E, Marsh P, Staunton M. Credit Suisse Global Investment Returns Sourcebook. Zurich: Credit Suisse; 2014. Available from: <http://doc.xueqiu.com/14cdbae48e74653fe7546fe0.pdf>
6. Bengen WP. Determining withdrawal rates using historical data. *J Financ Plan*. 1994;7(1):171–180.
7. Cooley PL, Hubbard CM, Walz DT. A comparative analysis of retirement portfolio success rates: Simulation versus overlapping periods. *Financ Serv Rev*. 2003;12:115–128.
8. Scott JS, Sharpe WF, Watson JG. The 4% rule – at what price? *J Investment Manage*. 2009;7(3):31–48.



## Casting the nets wide in the quest for ways to 'Africanise' research and teaching

### BOOK TITLE:

Africa-centred knowledges:  
Crossing fields and worlds



### EDITORS:

Brenda Cooper and Robert Morrell

### ISBN:

9781847010957 (hardcover)

### PUBLISHER:

James Currey, London; GBP50

### PUBLISHED:

2014

### REVIEWER:

Menán du Plessis

### EMAIL:

Menan.du.Plessis@gmail.com

### AFFILIATION:

Department of  
General Linguistics,  
Stellenbosch University,  
Stellenbosch, South Africa

### POSTAL ADDRESS:

Department of  
General Linguistics,  
Stellenbosch University,  
Matieland 7602, South Africa

### HOW TO CITE:

Du Plessis M. Casting the  
nets wide in the quest for  
ways to 'Africanise' research  
and teaching. *S Afr J Sci.*  
2015;112(1/2), Art. #a0144,  
2 pages. [http://dx.doi.  
org/10.17159/sajs.2016/a0144](http://dx.doi.org/10.17159/sajs.2016/a0144)

A book entitled '*Africa-Centred Knowledges*' is bound to be opened with hope, arriving as it does at a time when university leaders throughout South Africa are reflecting urgently on the state of their institutions, and when civil society as a whole is feeling grave concern about the slow pace of transformation and Africanisation in higher education. Where academic transformation involves material issues – such as greater student access and support, as well as the right of workers on the country's campuses to decent jobs, fair pay and benefits – these may now have been temporarily, if belatedly, addressed. However, the concept of Africanisation, even while it is intertwined with transformation, is harder to pin down – and it is this topic that the papers collected in this book bravely grapple with.

The book comes with a reassuring pedigree: we are told that it had its origins in an academic collaboration spanning three continents, which began in 2005 with footholds in departments of African Studies at Brown University in the USA, at the University of the West Indies at Mona, and at the University of Cape Town in South Africa. As one of the outcomes of this collaboration, the Africa Knowledge Project was established in 2009 at the University of Cape Town, and it was under its auspices that a workshop was held in 2011, where the papers collected here were first presented for discussion and refinement.

The book is split into two parts: the first contains papers of a broadly theorising nature, and the second contains case studies that demonstrate various practical aspects of the theoretical concepts. The introductory chapter by Brenda Cooper and Robert Morrell (p. 1–20) provides a useful overview of the kind of thinking that informs the book as a whole, and is commendably characterised by a desire to avoid simplistic binary oppositions. For example, they tell us that the collection 'is poised in a creative and productive third space between the polarisation of the bad place of Eurocentrism and the immovable rock of Afrocentrism'. We are assured that all the contributors appreciate that 'Africa is not homogenous or monolithic'; and we are also assured, if only obliquely, that while all the authors 'consciously contribute' to 'southern theory', they are aware that some countries wrestling with colonial legacies and enduring social inequalities are physically located in the northern hemisphere. Lastly, for those of us out of touch with trending topics in this area of rapid academic growth, there is a brief account of the 'maelstrom surrounding the definition of modernity'. Amongst other things, we are told that 'western capitalist forms of the industrial nation state, a market economy and, most especially, positivistic scientific methods, have come to define the substance of modernity itself'.

The old-fashioned reference to scientific positivism is probably meant to arrest our attention, with its connotations of vintage debates around absolute truth and the role of empiricism in the social sciences. It also prompts us to reflect on the book's title, with its awkward (some might say pompous) use of 'knowledges' in the plural. This is no doubt equally deliberate, and drives home the well-established idea that knowledge is not monolithic, and that there is scope for pluralism, at least within the social sciences and the humanities. Most of us will have no difficulty in conceding the relative and not always empirical nature of knowledge in such domains. A first small discomfort may be felt, however, when it starts to emerge that the agenda of some contributors to the book is to conflate this pluralistic conception of 'knowledge-making' in the humanities, and the concept of an *Africa*-centred approach – in short, by suggesting that African ways of creating, formulating and disseminating that which is known about society and nature are somehow 'other'.

It is clear that a dilemma arises when we try to address aspects of 'the African predicament', such as the extreme social divides still so deeply entrenched in many of its countries, and associated with these, the almost brutal impediments to access, whether it might be to education, healthcare, housing, welfare, or economic participation. The dilemma is that when we try to confront such difficult things, it is all too easy to speak in terms that suggest the need to 'catch up' with 'more advanced' societies located in a notional 'northern hemisphere' – and by using such terms, to imply that Africa (that rather vast continent) is entirely lacking in history, innovation, systems of knowledge, vision, strengths and capacities of its own.

The contributors to the book are deeply mindful of this problem. Their approaches are diverse and fairly complex, but the elaborate frameworks they each build for their particular purposes are all loosely woven through with one or more of the various strands alluded to above. The catch that is hauled out by means of these different fishing traps is intriguing, if in the end disturbing. While the bulk of it seems to be made up of fishes that could have been caught almost anywhere in the world, there are on the other hand a few specimens so narrowly localised in distribution that they hardly qualify as universally African. Lastly, there are a few creatures that are fascinating in their own way, yet perhaps are not really part of the intended catch at all.

A valuable contribution to the first part of the book is Lesley Green's proposal (p. 36–50) for some rethinking on the subject of indigenous knowledge, which has become such a fashionable and reified notion that it is often referred to simply as 'IK' (and which, of course, is hardly a topic of uniquely African or even 'southern' relevance). Particularly important, even though it is perhaps of only local relevance, is the author's call for a re-examining of the notion 'that "indigenous knowledge" exists primarily as a static inheritance with potential for wealth creation in the knowledge economy.'

Another interesting contribution in this section is the examination by Mbugua wa Mungai (p. 93–108) of ways in which popular musicians in Kenya may use their art to express subversive critiques of society and politicians. While it is not clear why this should be seen as a specifically African phenomenon, the paper itself is richly informative.

© 2016. The Author(s).  
Published under a Creative  
Commons Attribution Licence.

Other papers in the first part include a contribution by Lansana Keita (p. 23–35) that confronts myths about Africa, and which interestingly seems to go against the grain of the rest of the book, in its suggestion that 'contrary to those who argue for a relativistic approach to knowledge, it is indeed possible to establish empirically certifiable knowledge, even though all knowledge exists in a social context'. A paper by Signe Arnfred (p. 51–63) considers gender issues in the context of 'development discourse', and discusses some of the ambiguities and north/south contradictions in the 'discursive struggle' process, as well as the 'process of co-optation by the World Bank'. (A further paper on the topic of gender issues (p. 163–177) is assigned a place in the second part of the book, perhaps because it focuses specifically on Ghana. The authors, Akosua Adomako Ampofo and Michael P.K. Okyerefo, examine the ambiguous role played by male religious figures in promoting gender equality in that country.)

The two final papers in the first section are literary analyses. The contributor of one of these (p. 64–77) draws on evidence from 'the African novel' – in fact the work of Chinua Achebe, Dambudzo Marechera and Ben Okri – to conclude that African time is 'non-linear', and reflects a different kind of knowledge, where time is apparently 'circular', 'layered' or 'folded'. The author of the other literary paper (p. 78–92) examines the creative work of Helen Oyeyemi, and concludes that this writer has the (apparently remarkable and uniquely African) capacity to perceive the world in ways that go beyond the preconceived. Each of these conclusions in its own way deepens the small sense of unease some readers will have begun to feel from the outset.

The second part of the book contains papers that are intended to illustrate various aspects of the 'Africa-centred knowledges' mooted in the first part, as they might be manifested in practice. Several, however, merely suggest a need – surely well established and long recognised internationally – for approaches to information-gathering and design to be less top-down, and to involve target communities in a more egalitarian way. One paper (p. 111–125) speaks, for example, about the scientific management of fish resources in Namibia, and tells us that 'fishing captains have a deep knowledge of the fish they hunt', and that 'any decisions made about fishing in Namibia would benefit from including fishers' knowledge, and from a conversation between scientists and fishers that develops a holistic and shared eco-knowledge'. Another paper in this category (p. 126–141) looks at the roll-out of information communication technologies in southern Africa, and some of the unexpected problems, aside from purely technical and logistical ones, that have been encountered along the way. The authors conclude that the design of applications and user-interfaces could be improved by a greater awareness of 'local-context information', and by communicating and interacting more directly with local end-users. A third paper in this category (p. 142–162) addresses urban planning and the making of knowledge about health and environment, with specific focus on the problem of housing provision in Cape Town. The authors find that amongst the residents of Khayelitsha, the concept of well-being includes a spiritual dimension. This is apparently so unusual (and so African) that they are led to conclude that 'northern urban-planning approaches and southern urban realities do not make a perfect match'.

Lastly, two papers look at approaches to learning, researching and writing in the academic environment itself. One of these (p. 178–190) is a 'meta-reflection' on processes that went into the production of a previous book on postgraduate writing, which addressed problems encountered by young researchers from 'the geopolitical periphery'. The paper speaks to the difficulties faced by emerging researchers who are expected to present their findings in a language that might be only their second or third, in a register deemed appropriate for a particular discipline – where these dominant conventions are seen as a further manifestation of 'northern' values. The authors describe their pedagogic experiments with writer circles, in which postgraduate students could safely try out the formulation of their ideas 'before taking them in writing to more threatening audiences'.

In this kind of context, they suggest, the taking of risks may have positive benefits, and what is more, may be more easily accommodated in countries like South Africa 'than in the more stable "northern" institutions where quality-assurance regimes are more deeply entrenched'. It is not clear, though, why it should be assumed that similar difficulties are not faced by millions of young people around the world today, or that similar student-centred approaches to learning, teaching and evaluation have not been tried out and implemented at higher institutions around the world.

The second paper in this last category (p. 191–205) provides a case study by Leadus Madzimu of a Zimbabwean secondary school, and finds a link between the unexpected academic success of deeply disadvantaged learners – and the positive sense of community-located identity and self-worth that is inspired by the concept of *ubuntu*, or *hunhu* in the cognate Shona word used by the author. This principle is said to empower young students, in the sense that personal application and aspiration are supported as a collective good. What this paper reflects on are really some of the precursory factors to the successful making and acquisition of knowledge itself, but it is nevertheless one of the more intriguing specimens hauled up in the by-catch. It would be interesting to know about any cases in which the concept of *ubuntu/hunhu* might be present, and yet the expected success does not occur; and it would also be interesting to know whether similar concepts (having similar effects) exist under different names in the vast areas of Africa where other, unrelated languages are spoken.

On reaching the end of the book, it is difficult not to think that the promised expounding of 'Africa-centred knowledges' has turned out to be the big fish that got away. Such a troubling thought even inspires a re-reading of the opening chapter, in a last attempt to isolate what kind of knowledges the various contributors would have us understand as both uniquely African and Pan-African.

On second reading, though, the introduction is even more disconcerting than it seemed at the outset. For example, the authors add an enhancement of their own to the paper mentioned earlier, on 'the African novel' – remarking (p. 10) that its author 'thickens the interrogation of modernity' by scrutinising 'the fallacy of time as linear', and stating that he 'embeds Afro-modernity in a concept of time in which the past and the present are intertwined, destroying the binary between archaic pasts and progressive presents'. It seems we are to understand that there is an African concept of time, and that it is different from everyone else's.

They also tell us that 'embodied knowledge' represents a source of evidence too often discounted by mainstream gatekeepers, and they point us to the paper on urban planning, whose authors, they suggest (p.16) 'derive a visceral sense of the relationship between people and the cities they inhabit, by inviting them to map their bodies onto their living spaces'. It seems we are to understand that there is a viscerally accessible level of knowledge that is uniquely African, and which is oppressively disregarded by those who take a 'positivist' approach, or those with a 'northern' orientation in general.

Finally they suggest (p. 18) that one of the things constitutive of Africa-centred knowledges is a 'hospitality towards all ideas'. It is not clear why a liberal and open-minded approach to ideas should be presented as uniquely African. The very assertion leaves us with a renewed sense that a few (but by no means all) of the contributors to this project may be attempting consciously or unconsciously to seek validation and acceptance of various treasured approaches of their own – by projecting them as somehow special, and somehow African.

This book was opened up with eagerness, in the hope that it might help to formulate and illustrate some of the changes so direly needed if South African universities are to start achieving full transformation, and Africanisation in particular. If this is all that is on offer – almost a quarter of a century since it became clear that the end of apartheid was in sight – we should probably brace ourselves for another year of troubles and disruptions on our campuses.

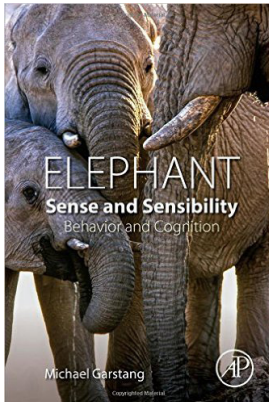




## What separates them from us?

### BOOK TITLE:

Elephant sense and sensibility:  
Behavior and cognition



### AUTHOR:

Michael Garstang

### ISBN:

9780128022177 (softcover)

### PUBLISHER:

Academic Press, London;  
USD42.46

### PUBLISHED:

2015

### REVIEWER:

Rob Slotow

### EMAIL:

slotow@ukzn.ac.za

### AFFILIATION:

Amarula Elephant Research  
Programme, School of Life  
Sciences, University of KwaZulu-  
Natal, Durban, South Africa

### POSTAL ADDRESS:

DVC Suite, 1st Floor Desmond  
Clarence Building, Howard  
College campus, University of  
KwaZulu-Natal, Durban 4041,  
South Africa

### HOW TO CITE:

Slotow R. What separates  
them from us? S Afr J Sci.  
2016;112(1/2), Art. #a0140,  
2 pages. [http://dx.doi.  
org/10.17159/sajs.2016/a0140](http://dx.doi.org/10.17159/sajs.2016/a0140)

© 2016. The Author(s).  
Published under a Creative  
Commons Attribution Licence.

Michael Garstang has provided us with a short treatment of elephant behaviour and cognition, which sets out to place aspects of elephant behaviour, psychology, and cognition into a biological context. It is certainly a challenging task, given that we actually know surprisingly little on this subject, and that there are major risks of anthropomorphising interpretations in this context. This risk is emphasised by the author in the introduction, in which he points out that 'while we are able to consider significant findings, these will not be without contention, not in many cases fully supported by definitive scientific evidence'. His approach is to explore the subject by describing both his own observations (he has studied sound communication in elephants extensively) as well as drawing on those from colleagues and the literature, and to then place these observations into a theoretical framing. In his own words 'we draw upon unverified anecdotal evidence not as proof of concept but as possible guidance to the development of a testable hypothesis or the formulation of a question'.

The book is relatively short, about 120 pages, and is structured into 14 (short) chapters, starting with a phylogenetic contextualisation followed by a description of the elephant brain and neural function. Garstang then proceeds with four chapters that deal with memory, morality, emotions, and empathy and altruism, largely, in his words 'in our attempt to penetrate the mind of an elephant'. These chapters are then followed by four chapters that deal with communication, language, intelligence, and teaching and learning. He finishes with a concluding chapter entitled 'Them and Us', which examines elephant conservation, management, and especially human conflict with elephants, including that through poaching.

The approach of the author is to develop a moral position about elephants as a species, and he poses the question at the start of Chapter 14 as 'at this point in our effort to penetrate an elephant's mind we need to ask what have we learned about an elephant's mind and what separates them from us?' As he moves through the various chapters, he continually references what he is saying about elephants back to human context, and he concludes on p. 104 that: 'Elephants are sentient beings differing from humans only in degree and not kind.' He indicates that this conclusion 'requires that we re-examine the relationship between ourselves and elephants' and poses the question of 'what sets us apart and what justifies the acceptance of significantly different privileges for one species versus another?'.

Throughout the book it is clear that the author has a particular agenda, to demonstrate that elephants are not that different from people, and that we have a moral obligation to treat them with due respect and concern, especially in the context of protecting them for persistence. Does this make the book one-sided, of lesser value, or of lesser interest? At the start of the review, I highlighted the challenge of anthropomorphising animals, and my review so far should emphasise to the reader that the author treads a delicate line throughout the book, the line of is it science or pseudoscience?

I have been deliberate in this review in laying out the challenge, and the reader may interpret that I do not believe that the book has value, or that the book is not worth reading. I have been deliberate in creating this context, but I now want to approach the assessment differently. One of the key issues that we have been tussling with in the conservation community in South Africa, and more broadly, is how we treat animals from a welfare perspective. This issue has been forefront in the debates on management of elephants in South Africa, and a very important context for conservation has emerged. As part of the process undertaken by the Department of Environmental Affairs and Tourism of developing norms and standards for the management of elephants<sup>1</sup>, all the aspects that are dealt with by Garstang were taken into consideration. A conclusion was reached that elephants are not simply like other animals. The norms and standards<sup>1</sup> articulate this conclusion in the following way: Clause 2(a)(vii) 'recognises their sentient nature, highly organised social structure and ability to communicate'; Clause 3(a), 'elephants are intelligent, have strong family bonds and operate within highly socialised groups and unnecessary disruption of these groups by human intervention should be minimised'; and Clause 3(h), 'management interventions must, wherever practicable, be based on scientific knowledge or management experience regarding elephant populations and must (i) take into account the social structure of elephants and (ii) be based on measures to avoid stress and disturbance to elephants'. In the summary for policymakers from the *Assessment of South African Elephant Management*, the authors point out that elephants exhibit characteristics 'in combination, and to such a degree, that people of many different cultures and backgrounds agree that elephants must be managed with a degree of respect greater than that afforded to most other species of wild animals'<sup>2(p.SPM3)</sup>. It is because of this recognition that elephants are managed within South Africa in a manner that is compatible with their complex psychology, which requires that we approach elephant management from an elevated status of welfare concerns.

While this was the sentiment expressed in the process and outcomes of the norms and standards in South Africa, and is dealt with in various chapters of the Assessment summarised above, it is very difficult to access a readable integration on which this sentiment is based. I see this as the main value of the book by Garstang. What he has done, is to systematically work through our understanding of elephant biology and psychology so that we develop an understanding that elephants are not simply other animals. An understanding that there is complexity and depth to elephants which reflects strongly our own biology, which has evolved from being a social, long-lived species with overlapping generations that have enabled evolution of cultural attributes such as language, and corresponding neurological development. The arguments he puts forward are a mixture of very little scientific evidence, anecdotes, observations, and conjecture, especially in creating parallels with humans. While there are challenges with this approach, to his credit he emphasises these upfront, and, I believe, does a credible job of pulling it all together. I believe that he will convince readers that elephants are indeed complex social species, which are different from other animals. The challenge will be whether he has convinced the reader strongly enough for us to change our collective moral viewpoint to recognise them as sentient, and different from us only in degree, and requiring an

assessment of the 'privileges' we accord them. Notwithstanding this, I do need to point out that the book is not comprehensively referenced, and that the author does not make enough reference to the various chapters in the Assessment, especially the one on ethics by Lötter et al.<sup>3</sup>

While many readers may find the book challenging in that it skirts, and sometimes crosses into, anthropomorphising, the book does a very good job of putting pieces of a puzzle together that enables us to understand that elephants are different from other animals, and maybe not that different from humans. Because the book is short, and because he makes the 'evidence' accessible, the book is an easy read. I believe that it provides an excellent resource for making us think more about what we are and what we do. This makes the book an important read for conservation practitioners, not only in South Africa, for undergraduate and postgraduate students who are developing their world view, and for academics who may be set in their ways, as a means to think outside the box. For anyone who wants to form and express views on how we should manage and engage with other species, the book provides a reference point, underpinned by substantive, if passionate, argument,

for a particular perspective; a perspective, I believe, that needs to be incorporated into the debate we should continually be having around conservation practice and management.

## References

1. Department of Environmental Affairs and Tourism. National norms and standards for the management of elephants in South Africa, in terms of the National Environmental Management: Biodiversity Act, 2004 (Act No. 10 of 2004). Government Gazette No. 30833, 29 February 2008, South Africa.
2. Scholes RJ, Carruthers J, Van Aarde R, Kerley GIH, Twine W, Grobler DG, et al. Summary for policymakers: Elephant management in South Africa. In: Scholes RJ, Mennell KG, editors. Assessment of South African elephant management. Johannesburg: Wits University Press; 2008. p. SPM1–20.
3. Lötter HPP, Henley M, Fakir S, Pickover M, Ramose M. Ethical considerations in elephant management. In: Scholes RJ, Mennell KG, editors. Assessment of South African elephant management. Johannesburg: Wits University Press; 2008. p. 307–338.

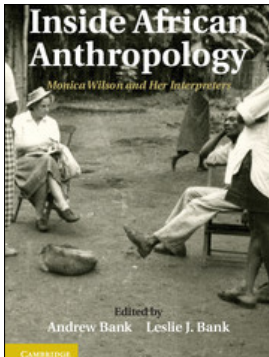




## Whose knowledge? The politics of scholarship in South(ern) Africa: A critical review of *Inside African Anthropology*

**BOOK TITLE:**

Inside African anthropology:  
Monica Wilson and  
her interpreters



**EDITORS:**

Andrew Bank and Leslie J. Bank

**ISBN:**

9781107029385 (hardcover)

**PUBLISHER:**

Cambridge University Press,  
Cambridge; USD99.99

**PUBLISHED:**

2013

**REVIEWER:**

Shahid Vawda

**EMAIL:**

Shahid.vawda@wits.ac.za

**AFFILIATION:**

School of Social Science,  
University of the Witwatersrand,  
Johannesburg, South Africa

**POSTAL ADDRESS:**

School of Social Science,  
University of the Witwatersrand,  
Private Bag 3, Wits 2050,  
South Africa

**HOW TO CITE:**

Whose knowledge? The politics  
of scholarship in South(ern)  
Africa: A critical review of  
*Inside African Anthropology*.  
S Afr J Sci. 2016;112(1/2), Art.  
#a0142, 4 pages. [http://dx.doi.  
org/10.17159/sajs.2016/a0142](http://dx.doi.org/10.17159/sajs.2016/a0142)

In this review, I raise the critically important issue of scholarly knowledge production among a range of scientific disciplines, particularly those in both the humanities and the sciences that rely on extended periods of fieldwork, through a critical reflection on the book by Andrew and Leslie Bank on the impressive work of the late Professor Monica Wilson. The Bank brothers have put together an important and impressive collection of chapters devoted to two critical issues: in part an initial assessment of Monica Wilson's work over her entire career, but more critically, a perspective of her interpreters – her field assistants and students, as well as various political interlocutors, friends, colleagues and of course her family, particularly her father. The book traces a chronological, perhaps a genealogical, intellectual development of Monica Wilson, alongside that of her many co-workers, in the production of a series of books, journal publications, presentations of various kinds and talks of hers. It begins with her Cambridge days – where she converted from studying history to anthropology, through to her fieldwork in Pondoland and East London and then her years of fieldwork with her husband Godfrey Wilson in Tanzania, and later to her work in Lovedale, Fort Hare and Rhodes University and finally her move to the University of Cape Town. The key focus of the book is on her co-workers and co-producers of anthropological knowledge. In the southern African context, this book is a full-length, but by no means exhaustive, exploration of an important second-generation ancestor of the discipline of anthropology and its consolidation in universities such as Fort Hare, Rhodes and Cape Town. The book is in large measure based on the Monica and Godfrey Wilson archives at the University of Cape Town, the significance of which lies squarely in the intimate connection between the personal dimensions of ethnographic research and the institutionalisation of research at a university.

The book mainly devotes itself to the collaborations and engagements with her field assistants. These collaborations and engagements hinge on two central concepts: the 'insider' (and surprisingly less on its opposite the 'outsider') and 'interpreter'. The book also contributes to a now well-developed field of interest related to the fieldwork assistants, the language teachers and interpreters. In many instances they were the equivalent of the ethnographers. In what follows is a review based around these three concerns: being an insider, and by implication, an outsider; the relationship to interpreters; and the research and knowledge production relationship as a developing field of enquiry.

Monica Wilson was an insider as well as an outsider in her various personae. The insider-ness relates to her intimate knowledge of the Eastern Cape: where she lived and was educated (Chapter 1), where she imbibed the missionary zeal of her parents (Chapter 1) and where she undertook her first major fieldwork (Chapters 2 and 3). Some of these 'insider' issues that concern her own worldviews need much more amplification. For example, the relationship between the missionary vision to change the world, albeit to a Christian one, with which she grew up and its relationship to her anthropology are important. By this I mean not the superficial idea that there was this influence – it is unmistakable – but how did that idea of missionary-induced social change translate or mutate over time and imbricate itself in her work. This question is important because there is a literature in South Africa that is not sympathetic to missionaries. And indeed this sentiment is captured in the often repeated phrase: 'When the missionaries came they had the Bible and the land was ours: now we have the Bible and no land.' That Wilson in many ways symbolises the acceptance of a benign missionary influence, needs much greater interrogation than the book gives.

As a developing fieldworker, despite her intimate knowledge of the Eastern Cape, she was an outsider. There were limits not only as a female anthropologist in the 1920s and 1930s that she faced both at Cambridge and in South Africa, but also as a fieldworker who required the intervention of a range of people, such as her father, other missionaries, priests, shopkeepers and their wives, mission educated English-speaking African clerks and quite importantly black politicians, such as Dr Walter Rubusana and Clements Kadalie, to ease her way into fieldwork and to protect her. Clearly her relationship with her father and his network was vital to her initial fieldwork, notwithstanding her own actual field research.

She was very much the insider who chose to stay in South Africa to continue her academic career here, within the confines of racial discrimination and its institutionalisation in apartheid, and to support and nurture her students and field assistants to carve out their own academic and other careers (Chapters 6, 7 and 8). That she chose to do so was a political and personally courageous decision of significance: a commitment to advance a scholarly enterprise in a country well known for strong streaks of anti-intellectualism mixed into a cocktail of racial prejudice, ethnocentrism and white superiority. Indeed, she was well aware of the way in which racial prejudice and apartheid marginalised and discriminated against her black students and colleagues and her insider fights against universities' acts of discrimination and prejudice, in particular the University of Cape Town, are well known. That she supported her black students and colleagues and engaged in research and debates with them earnestly was in contrast to some of her white colleagues and speaks volumes of her commitment to the development of a black intellectual class. This insider-ness is highlighted sharply in relation to Archie Mafeje (Chapter 8), Leonard Mwaishumo (Chapter 5), Godfrey Pitje (Chapter 6) and Livingstone Mqotsi (Chapter 7). Yet the list of those of her white students that she trained brings into sharp focus the uneven and structurally unequal access to universities, research resources and the possibility of publication and the proportionately different scholarly success: Berthold Pauw; Max Marwick; Peter Carstens; Peter Rigby; Jean Comaroff; John Comaroff; Colin Murray; Martin West; and Pamela Reynolds (p. 5), all of whom are, or were (some have passed on), highly respected

academics at well-known universities both here in South Africa and internationally. Yet the question may be asked as to who were the insiders and outsiders? The white students that passed through Wilson's lectures and supervision were also insiders of a kind, benefitting immensely from the insider privileged position of their class and race. Mafeje, Mwaisumo, Pitje and Mqotsi might well be insiders of African anthropology, a small coterie of Wilson's well-known black research assistants and students that pales by comparison. The contrast is dramatic and the insider politics of not just southern African anthropology, but of every scientific discipline at South African universities needs to be explored, which, surely, is the implication of quoting Nancy Jacobs' work on the research assistants in the field of ornithology. This brave book begins that journey of exploration and documentation for anthropology, as Andrew Bank states, in departure from the 'official' history. It should be the case for every scientific discipline in South Africa.

As a seasoned academic within the rapidly changing political environment in the 1950s to 1970s, Wilson confronted the narrowing alternative spaces that she and public intellectuals like her could take advantage of in the gathering menace of apartheid. This is not to question her personal, intellectual and academic integrity: she rejected attempts by the Institute of Race Relations to recruit her as a spy on her assistants; she provided sustained and persistent support for academics such as Mqotsi and Mafeje despite the seeming hopelessness in the face of state opposition and brutality and the cowering stance of universities in the face of government pressure (Chapters 7 and 8). It is rather a context of the constraints and limits she faced and over which she had little control but she fought against and provided what space there was to continue, and perhaps even expand such scholarship. One may question if being an outsider under such circumstances was a disadvantage? Some theorists have argued that the marginal or liminal position of the outsider or stranger is an advantage. What then might such an outsider position have been to Wilson's favour and that of her interlocutors? The possibility of being both an insider and an outsider in relation to what she could and could not do for her students and colleagues is clearly demonstrated in the book, but needs greater amplification as an example of the political economy she had to contend with and the severe constraints she operated within in the institution of the university. This is well beyond the 'personal' and 'experiential' motifs that guide the substance of the book's chapters. Probably further research in the Wilson archives is required and perhaps a re-assessment of the conceptual tools to interrogate those archives. Such research, as I have suggested above, is vital to the South African intellectual enterprise going into the future.

The definition of the word 'interpreters' and characterisation of the political and social context in which Wilson and her interpreters worked as variously colonial or neo-colonial, as well as segregation and apartheid are relevant. To quote Leslie Bank:

*Our use of the concept 'the interpreter' takes its cue from a public lecture that Monica delivered in Grahamstown in 1972, the year before her retirement. She began by pointing out that the first interpreters in the southern region of our continent were Africans, not colonists. These were the bilingual Christian converts who had learnt literacy and communicative skills on mission stations and then worked as translators. (p. 7)*

This quotation has a number of reference points in the entire book. However, I draw attention to it by way of a few examples. Suppose we treat the Wilson archive as the anthropologist Anne Stoler<sup>1</sup> has done the Dutch colonial archives: to be read against and along the grain, as an ethnographic fieldwork site of inquiry. Clearly, if this archive was an ethnographic study of anthropologists going about ordinary everyday scholarly practices, here is the collective ethnographer (editors and some of their co-writers) adopting the emic position of the 'natives', that is to say the analyst adopting somewhat uncritically the position of the informant. This is a thread in the book where the editors treat Wilson very much in awe, that is to say a little less critically than might otherwise be the case. While I do not want to take away from the fact

that her interpreters were in fact bilingual Christian converts, I do want to raise three issues that arise from this perspective. Firstly, that the first interpreters were indeed indigenous or non-Europeans. But they may not have been all Christian converts. If we take a conventional date for the beginning of colonialism, the Dutch settlement at the coast in 1652, the first interpreters were certainly not Christian converts. Secondly, it is quite clear from all the chapters that in Wilson's writings there is the almost unstated presumption of an undifferentiated category of 'pagan' in opposition to Christian, or at least the point of departure from which modernity, symbolised by Christianity, is measured. And being Christian is the elemental criterion of positive social change. Even as Leslie Bank suggests she was more interested in actual social groupings than social categories (p. 195), these presumptive ideas are apparent throughout the book and perhaps this is an underlying assumption in her oeuvre. Certainly, a public lecture a year before her retirement in which she foregrounds Christian converts suggests that this is the case. Perhaps more research in the archive is needed to test this hypothesis. Thirdly, by extension and implication, the categorisation of people into exclusive groupings: Christian/non-Christian; European/non-European; colonialist/non-colonialist; pristine traditional 'tribes'/modernised groups. Such indeed may have been the kind of dominant binaries that were pervasive for most of the 20th century social science, and which Wilson incorporated into her oeuvre, but such exclusive categorisation of people – an analytical separation of reality – does not prevent actual hybridisation and the way in which the ideas and actions of Christian missionaries, Hindu ascetics, liberals, socialists, communists, nationalists of various sorts, Muslim clerics, Jewish migrants and refugees, amongst a host of other influences, have co-mingled creating a rich mosaic of ideas and their variable reception among people in southern Africa. Such categorisations as Wilson operated with tend to privilege a Western Christian missionary perspective and silence and make hidden the multiple acts of engagement, translation, interpretation and mediation that make it possible and often necessary to act in the world of the slaves, proletarianisation in the 20th century or that of immiseration and vast and deepening inequalities in the 21st century. That she used Christian converts, or perhaps more accurately mission-educated African assistants, consistently in all her fieldwork sites (in South Africa and Tanzania) who also occupied a particular segment or class position that distinguished them from the 'poor and dirty' (p. 86) lends another, and in my view, significant perspective on her ethnography. Indeed, the meta-theoretical frameworks with which she used, whether of the 'one history', 'culture contact' or 'single economy' variety, did not pay much attention to hybridisation/creolisation or transnationalism or indeed to the question of class, income and racial hierarchy – aspects that would have occupied the minds of Archie Mafeje and Livingston Mqotsi. One quite startling example on the question of racial and class hierarchy is the questionable acknowledgements by presenting her store trader hostess and key informant as a 'white woman of Scottish ancestry' (a vague label that might well apply to Wilson herself) in her first monograph *Reaction to Conquest* while her notes and PhD thesis have her as a woman of 'coloured blood'. I am never sure what to make of these kind of remarks that clearly exist in the minds, words and reactions of many people, apart from its racist connotation, and that the social category of people labelled 'coloured' are usually considered a result of an ahistorical biological 'mixing of blood', a popular imagined accidental miscegenation, whereas the other social or racial categories in South Africa are defined and privileged in various ways by their historical, cultural, ethnic, language, national or geographical origins. It is perhaps best that we learn from Wilson's logical and empirical inconsistencies on these matters and come to terms with our hybridised or creolised class-based past and present.

Apart from these historical omissions and theoretical lacunae, there is the question of why she shifted in her acknowledgements of her field interpreters over time, from the mere acknowledgement in the co-production of important ethnographic insights to her openly stating towards the end of her active career where her debts lay. Borrowing from James Clifford<sup>2</sup>, Leslie Bank argues that the conventions of the 1930s ethnographic authority – the immersed and accomplished linguistically agile ethnographer whose knowledge and authority of the people studied

is virtually unquestioned – dictated the acknowledgment but also forcefully the exclusion of her debt to fieldworkers. It seems to me that Wilson later in life acknowledged to her credit the error of this thinking. And in this she was in advance of anthropologists in the rest of the world.

Further, while it is well known that she was an amazing teacher, superb supervisor, and hard-nosed disciplinarian on matters of rigorous research (Chapter 10), she was equally a patient listener to her students and research assistants and always willing to engage in debate (Chapter 6). That we should have more scholars like this in South Africa shouldn't be debatable. Wilson provides the model of what an excellent or ideal university scholar should be. It is worthy to note she was less patient with inefficient university bureaucrats, and here one can only speculate on how she would, albeit in her retirement period, have received and reacted to Professors Ramphele and West, anthropologists and VC and DVC, respectively, at the University of Cape Town in the recent past.

Yet there is uneasiness with this comfortable vision of the ideal scholar and mentor, that Wilson provided the intellectual stimulus for that small group of black anthropologists who learnt their craft from her. Mafeje, Mqotsi and others shared the intellectual tradition of the Unity Movement which would have steeped them in their rich tradition of reading not just socialist classics and that of their leadership such as Isaac Bangani Tabata (also a Lovedale student in the 1920s) whose writings also describe the conditions of peasants and workers in the Eastern Cape, and Benjamin Kies, but others such as Benjamin Farrington, whose work on 'classical' Greece and Rome was as much a class analysis of traditional societies, and Nosipho Majeke (Dora Taylor) on the ambiguous role of missionaries in conquest. Of course it is still to be explored and explained why and how it is that these insights were or were not applied to their ethnographic work in South Africa. Was this the Marxism that Wilson objected to in her engagements with Mqotsi and Mafeje? More importantly, is there a gap between the politics of people such as Mqotsi and Mafeje with the Unity Movement background and their anthropology inspired by Wilson? Is there a kind of theoretical and ideological dissonance or schizophrenia between the reception of Europeanised socialist theories and anthropological theory? Is there a nativism aspect to 'insider anthropology'? Are there unresolved contradictory perspectives between a fluid modernity embedded in the socialist project and a kind of anthropology that wished to capture the 'tradition' before it disappears? Wilson's anthropology was scholarly and political and probably missionary inspired, but her politics operated without deep resonance and organisational reach to the people she studied. Did Mafeje, Mqotsi and Pitje and others have this luxury, given their involvement in liberation politics? Did their muted and extinguished academic careers in South Africa force upon them lesser alternatives? Can one really speak of an insider anthropology which always seemed to create faltering pathways and eventually cut short black academic life?

I wish to point to where Leslie Bank questions the concept of hidden colonial context of anthropology to examine Wilson's work and her interpreters. He in particular rejects Sanjek's<sup>3</sup> formulation of the relationship between anthropologists and their field assistants as 'hidden colonialism'. He suggests that such a formulation is inadequate and rather shifts the focus to the 'experiential' (p. 14), to 'the view from the tent'<sup>4</sup> on the grounds that it provides for an assessment of the agency of such interpreters. Indeed Bank places considerable value in using Clifford's distinction between three kinds of ethnographic writing – inscription, transcription and description (p. 125) – to emphasise the active role of Wilson's interpreters. While this provides illumination for Monica Wilson's writing or the way in which she produced her monographs and journal publications, it assumes that there is a university or research institution where these kinds of writing culture or society can take place for a group of emerging black anthropologists. I would suggest that for anthropologists such as Mqotsi this was more absent than possible, and that the very university itself, as a learning and research institution in South Africa, can be an alienating place for black scholars. In this sense, the insiders of African anthropology are very much the outsiders in the formal institutional arrangements of tertiary education and research.

The exploration of the relationship between university trained anthropologists and their various assistants is not particularly new in the discipline of anthropology. At least since the 1980s, Stocking<sup>5</sup>, Sanjek<sup>6</sup> and Clifford<sup>7</sup>, among others, have been focusing on the relationship between anthropologists and their assistants, and in doing so have explored a number of different dimensions of this relationship, including the very important question of how anthropologists relate to the people they study, how they access information and generate data and how such data are written up. Ethnography, or more correctly ethnographic writing, consists of a range of research methods and techniques; written down observations; headnotes and field notes; anthropologists' journals; diaries; surveys that were undertaken; interviews; reflections on the personalities of fieldworkers and the particular predilection for a set of methods that are consistently used; the ways in which the observation, the interview and the act (and the kind or type) of writing takes place; the ways in which the observation and the interview are being observed by 'the natives'; the break/rupture or disrapture of the everyday practices; the act of writing field notes and so on. I have indicated how Clifford<sup>2,3</sup> has distilled these insights into three kinds of writing and how that has been fruitfully used to illuminate Wilson's writing. Others have raised a host of other issues that have now been explored and analysed around the mystic and 'authenticity' of fieldwork, or participant observation, and the very act of being an influence in the lives of people has come under scrutiny. Indeed the very presence of an anthropologist, despite the myth of not disturbing the everyday, or even a significant ritual or celebratory event, is in fact an introduction of social change. For example, a refusal to consume food (the anthropologist has food taboos which need to be explained); or does one turn away when violence against a child or woman takes place? Does one protect a would-be victim of rape while engaged in participant-observation? How indeed does one react to police brutally suppressing a protest against low wages, or demands for legitimate political or social rights? All of these issues are now painstakingly dissected in the anthropological literature, and in many cases are cast in various posturing, theoretical, methodological, political and seriously personal, moral and ethical dimensions. Indeed, in many cases ethnographic research now raises serious ethical issues in university research committees that consider research proposals, as Reynolds observes in her chapter, and wonders whether Wilson's research proposals, as it were, would muster a nod of approval from a university research ethics committee today (Chapter 10, p. 318).

In South Africa, this kind of reflection on the ethnographic authority, least of all on fieldworkers and their contributions, has hardly begun, although there have been attempts in the past to point to the insider aspects of anthropology such as Lekgoathi's<sup>8,9</sup> work on Isaac Schapera's fieldwork among the Tswana.

In some respects, only an inkling of these kinds of issues had been vaguely anticipated in the debate between volkekunder and liberal/radical anthropology in the 1980–1990s in an axis that seemingly bifurcated anthropology in this country. Wilson's insider anthropology as it is described in this book is not an exposé anthropology – it is not the high moral colouration and pedantic outrage of the Marxist inspired anthropology of the 1980s. It exposed rapid changes in people's lives through detailed description, but did not make an obvious posture. What might be the reasons for this? Perhaps a deep concern to describe accurately and articulate a scientific view – it was her belief in fieldwork, the truth of fieldwork, and its enduring value. The enduring value of her anthropology has become the historical and ethnographic record of necessity in any research in the Eastern Cape or Tanzania, rather than the record of theoretically inspired politically correct research. Of course we now know that even fieldwork is structured and informed by a range of issues – epistemological, political and personal, and in South Africa colonialism, racial segregation, discrimination and apartheid was central. For the South African academy this points to the lingering criticism levelled at anthropology (as it should be for any scientific discipline) as a colonial discipline, rather than one which currently questions the assumptions upon which we operate in our organised forms of social life in our actions, norms, representations and institutions (apart from philosophy – but that is another issue).

But this book should stimulate much more examination of anthropology and its history, particularly that of the field assistants. At least we know that the ethnographers in the Native Affairs Department not only used bilingual local elites such as teachers to gather information through questionnaires, but also offered their own interpretations. That both Wilson and the Native Affairs Department relied on local elites to do their fieldwork and carry out surveys and interviews says much about the less than ideal Malinowskian-type ethnographer in South Africa.

Despite my criticisms of the book, *Inside African Anthropology* is an extremely important book because it brings to light the once 'hidden' relationship between Monica Wilson and the field assistants and some of the students she had. It also serves as a model for which other disciplines (physical, health and humanities sciences) need to explore their own complicities. The book places emphasis on the 'experiential' – the practice of fieldwork or ethnography, and the downplaying of the structural context. Such an approach shows the complexity of these relationships and by implication shuns a simplistic notion that it was a one-dimensional 'hidden form of colonialism', or that it provided in an equally simplistic way the space for an indigenous intelligentsia to emerge. Indeed, these relationships were far from being equal: they were also often intimate, personal and fruitful, but that should not negate that they were probably also paternalistic, dependent, sometimes strained and difficult. Monica Wilson might have thought they were among equals as scholars, but these relationships were riven with the hierarchies of race and racial prejudice that emerged daily and which in its exteriority and interiority, outside of anthropology and inside the discipline within the university, prevailed and continues to prevail in its various forms. This is not to say that Monica Wilson was a racist, certainly not in crude and obvious ways, nor implicitly so, but she was situated in that structural and contextual racism that Max Gluckman<sup>10,11</sup> so brilliantly

exposed in his 'situational analysis' and which we still find ourselves embroiled in today.

## References

1. Stoler A. *Along the archival grain: Epistemic anxieties and colonial common sense*. Princeton, NJ: Princeton University Press; 2010.
2. Clifford J. Notes on (field) notes. In: Sanjek R, editor. *Fieldnotes: The making of anthropology*. Ithaca, NY: Cornell University Press; 1990.
3. Sanjek R. Anthropology's hidden colonialism: Assistants and their ethnographers. *Anthropol Today*. 1993;9(2):13–18. <http://dx.doi.org/10.2307/2783170>.
4. Schumaker L. A tent with a view: Colonial officers, anthropologists and the making of the field in Northern Rhodesia, 1937-1960. *Orisis*. 1996;11:237–258.
5. Stocking G. *Observers observed: Essays on ethnographic fieldwork*. Madison, WI: University of Wisconsin Press; 1983.
6. Sanjek R, editor. *Fieldnotes: The making of anthropology*. Ithaca, NY: Cornell University Press; 1990.
7. Clifford J. *The predicament of culture: Twentieth century ethnography, literature and art*. Cambridge, MA: Harvard University Press; 1988.
8. Lekgoathi SP. Colonial experts, local interlocutors, informants and the making of the archive of the "Transvaal Ndebele", 1930-1989. *J Afr Hist*. 2009;50(1):61–80. <http://dx.doi.org/10.1017/S0021853708003976>
9. Lekgoathi SP. Sophonia Poonyane and the researchers of Isaac Shapera in Colonial Botswana. 2010 [unpublished conference paper].
10. Gluckman M. Analysis of a social situation in Zululand. *Bantu Stud*. 1940;14(1–3):147–174.
11. Gluckman M. Some processes of social change illustrated from Zululand. *Afr Stud*. 1942;1:243–260.





## Participatory democracy for our time

### BOOK TITLE:

Are South Africans free?



### AUTHOR:

Lawrence Hamilton

### ISBN:

9781472534613 (softcover)

### PUBLISHER:

Bloomsbury, London; GBP16.99

### PUBLISHED:

2014

### REVIEWER:

Anthony Egan

### EMAIL:

antheagan2@gmail.com

### AFFILIATION:

Steve Biko Centre for Bioethics,  
University of the Witwatersrand,  
Johannesburg, South Africa

### POSTAL ADDRESS:

Steve Biko Centre for  
Bioethics, University of the  
Witwatersrand, Private Bag 3,  
Johannesburg 2050,  
South Africa

### HOW TO CITE:

Egan A. Participatory  
democracy for our time. S Afr  
J Sci. 2016;112(1/2), Art.  
#a0137, 2 pages. [http://dx.doi.  
org/10.17159/sajs.2016/a0137](http://dx.doi.org/10.17159/sajs.2016/a0137)

As I read University of the Witwatersrand's political philosopher Lawrence Hamilton's challenging new book, I could not but recall another work written in different historical circumstances but with a similar sense of urgency – *The Eye of the Needle* by Rick Turner<sup>1</sup>. Turner's book was first published in 1972 as a contribution to, indeed a radical critique of, the Christian Institute and SPRO-CAS (Study Project for Christianity in Apartheid Society) attempt to imagine a post-apartheid society. Although Turner had worked with SPRO-CAS, he was critical of its overtly liberal democratic conclusions, which he felt failed to understand the prerequisites for a new democracy. Given the nature of South African society then, Turner saw the only moral and practical solution in participatory democracy and socialism.

While not necessarily socialist as Turner was in 1972, in so many ways Hamilton is doing the same for South African democracy today – challenging our presuppositions, calling us out of our temptation to cosy consensus or despair. He starts with a shocking suggestion: the idea of freedom, enshrined in the Constitution and Bill of Rights, has simply not been realised more than 20 years after the transition to democracy. He argues:

*If freedom is conceived of in realistic, substantive and practically meaningful terms as 'freedom as power through political representation', we have no option but to conclude that South Africans are not yet free... (p.124),*

because in social, economic and political domains South Africans lack power. This provocative claim, which may seem to the complacent to be counterintuitive, is based on what he sees as the elite compromise (mainly between the African National Congress (ANC), the National Party and local and international economic interests) that failed to empower – to free – South African citizens (a similar view can be found in Habib<sup>2</sup>).

Two key elements of the transition 'pact' highlight this claim. The first can be summed up as the electoral system of party list-based proportional representation (PR), which placed too much power into the hands of parties and their leaderships. Although it was argued in the 1990s that the PR system was fair (particularly to minority parties who might otherwise vanish under a Westminster 'first past the post' system) and easy to manage, even its proponents agree that the party list PR model has disadvantages.<sup>3</sup> Party list PR is customarily used in small countries with fairly homogenous populations, not a good start for South Africa! But the most notable problem is that by voting for a party, citizens have limited ability to directly choose representatives as such; parties choose candidates and voters choose the party. The result is that voters have little say over who specifically goes to parliament and who stays in parliament. They cannot express disapproval with individual parliamentarians without voting against a party they might support. The party is all, the citizen is vote fodder.

The problem can be further exacerbated by situations in which said parties are riven with factionalism, authoritarianism or corruption, often resulting in direct manipulations of organs of the state (including those that exist to protect constitutionally guaranteed freedoms), and in situations in which one party may be dominant. Sadly, all these features are to be found to varying degrees within the ruling ANC. And many (including some from within the ruling party) see widespread manipulation within organs guaranteeing constitutional freedom. We might even say that we see a feature Southall<sup>4</sup> noted in many former liberation movements in government: the conflation of ruling party with government, state and nation.

The second problem with 1994 was the decision to embrace the globally dominant neo-liberal model of economy, which – despite affirmative action and black economic empowerment – substantially preserved the old order's economic status quo and weakened the basis for economic and social development. This decision was partly the result of a hope that by going with what was then considered the 'consensus', foreign investment would revive what was by the early 1990s a severely weakened economy. This has failed in the long term to deliver the promised goods. In addition, as Hamilton and many others note, the neo-liberal 'consensus' has had a very mixed history in delivering prosperity in developing economies (I am being generous in this observation) and is by no means the sole model for economic development even under a capitalist system. During the 1960s, for example, many Asian economies that became prosperous used a far more 'statist' and interventionist approach to development. One might add, as Cambridge economist Ha-Joon Chang has observed, that some of the most powerful free market economies (like the USA, China and Russia) have either risen to prosperity through capitalist forms of statism, or have continued to combine statism and neo-liberalism.<sup>5,6</sup>

To this arguably theoretical objection, Hamilton poses the important question: How has the 1994 decision affected South Africans' economic quality of life? Despite the clear emergence of a new black middle class, the majority of South Africans remain poor, with low levels of human development, particularly in what is considered a middle-income country. The reason for this situation is growing inequality, making us one of the most unequal societies in the world.

Moving beyond the specific politics of transition, Hamilton presents a philosophical account of why he embraces a definition of freedom as power. Freedom as power does not divorce liberty from the notions of power and representation. This is important because what he calls the liberal or republican model can easily become an abstraction rather than an expression of something concrete. Freedom as power does not reduce freedom merely to absence of external impediments, self-determination or absence of citizenship. It is not enough to be free from unjust constraint, from being the subject of a tyrant. Nor is it enough, Hamilton seems to be saying if I read him correctly, to have the right to vote every few years. Real freedom must entail active agency, participation in society and individual and communal development.

© 2016. The Author(s).  
Published under a Creative  
Commons Attribution Licence.



In short, Hamilton is advocating a form of what Turner called, four decades ago, participatory democracy, where individuals actively and directly choose who shall represent them, and when needed remove from office those who fail to deliver. Similarly, such participation must entail full and meaningful participation in an economic life. Under this account one can see why Hamilton presents the thesis that most South Africans are not yet free. The PR electoral system makes representation difficult and an economy that still excludes many from the possibility of fulfilment of their potential disempowers those who are poor.

Like Turner, Hamilton not only diagnoses the malaise but proposes certain remedies. He recommends a shift away from the party list PR system towards a mixed model, in which citizens directly elect persons as representatives *and* a party. The model that comes to mind (and to which he refers) is that of Germany, where parliament is constituted of both representatives chosen from a party list and directly elected members. It is an effective model that combines the advantages of PR and Westminster and reduces the disadvantages of both. (He also alludes in passing to Ireland, which is a somewhat different, more complicated system. Personally, having looked at both systems, I think Germany's model better meets the criterion of simplicity.) Hamilton also sees the need for the restoration of parliamentary sovereignty: in effect this means the reduction of one party dominance. He also, in a turn again reminiscent of Turner, sees the need for District Assemblies of citizens – participatory bodies marked by diversity that address policy issues. In a move that some may see as iconoclastic, he advocates plebiscite-based constitutional and policy revision. Similarly, he suggests that the demand for greater equality – a prerequisite of real freedom – may require revisiting the economic models and policies currently in place, because they simply do not work.

As I have already noted, this book, like Turner's *Eye of the Needle*, is challenging. When I first read Turner in the mid-1980s as an undergraduate, I was struck by how much of it was a product of its time, with an analysis of South Africa both prescient and problematic. The problem with it was that, with hindsight, some of its proposals were simply wrong: homeland leaders by the 1980s were no longer potential allies of the struggle, the homelands were no longer potential zones of freedom. Yet Turner's foundational point – the need for participatory democracy and the creation of effective grassroots organisations, alliances between groups sharing much in common, etc. – was right. I get a similar sense when I read this book.

From a historian's perspective, I do not share what I think is Hamilton's reading of the ruling party: a great party that has somehow recently lost its way and needs to find it again. Recent revisionist history suggests a very 'mixed' view of the ANC: allowing for difficult political circumstances on one hand, author bias and possible exaggeration on the other, Ellis<sup>7</sup> proposal of corruption and authoritarianism in the exiled ANC suggests a consistent shadow side to our major (but never sole) movement of liberation, something that might explain its less salubrious side today. The consolidation of the relatively more participatory democratic United

Democratic Front into the ANC during the transition period, with a relative eclipse of the latter in party leadership, might be a more nuanced perspective – which, however, in no way undermines the key points Hamilton makes, particularly his implicit unease with overt rejection of the ANC. Pragmatically, I would argue that the kind of change Hamilton seeks is unlikely to come from other parties, not least because they currently lack a strong power base, nor from the current ANC.

I also wonder whether the kind of *effective* economic reform, although manifestly necessary, is possible in the short to medium term. More than 20 years later, we may be faced with too few people with both skills and self-motivation to create a successful economy based on what would need to be done: redistribution of resources, creation of new jobs and the building up of enterprises among the poor. A bad education system can barely serve the unequal and unjust economy we have now. To make Hamilton's model work would entail real capacity-building on a massive scale, as well as a new work ethic. Both can be done, but I wonder whether the old and new elites have the will to do it. That said, if it is not done I can foresee more of the same ultimately self-destructive actions (service delivery protests, xenophobic struggles over scarce resources, etc.) that Hamilton rightly sees as a dangerous and mistaken short-term reaction to our highly unequal society.

These caveats (possibly an expression of a certain 'pessimism of the intellect') aside, I thoroughly recommend Lawrence Hamilton's book. If I may hark back one last time to Rick Turner, despite the problems I had with *Eye of the Needle* back in 1985, his is a book that at the risk of waxing sentimentally, profoundly affected my way of seeing my world at the time. Turner's broad vision, his underlying 'optimism of the will', captured my political and moral imagination. I have read it many times and still find it inspiring. In many respects this is what Lawrence Hamilton also does in this little volume. Faced with so many challenges, it is a tightly argued and passionately articulated call to rethink our society. And to act.

## References

1. Turner R. *The eye of the needle: Participatory democracy in South Africa*. Johannesburg: SPRO-CAS; 1972.
2. Habib A. *South Africa's suspended revolution: Hopes and prospects*. Johannesburg: Wits University Press; 2013.
3. Diamond L, Plattner MF, editors. *Electoral systems and democracy*. Baltimore, MD: Johns Hopkins University Press; 2006. p. 73–134.
4. Southall R. *Liberation movements in power: Party and state in southern Africa*. Pietermaritzburg: University of KwaZulu-Natal Press; 2013.
5. Chang H-J. *Bad Samaritans: The myth of free trade and the secret history of capitalism*. London: Bloomsbury; 2008.
6. Chang H-J. *23 Things they don't tell you about capitalism*. London: Penguin; 2010.
7. Ellis S. *External mission: The ANC in exile, 1960–1990*. Johannesburg: Jonathan Ball; 2012.



# ASSAf: Promoting scholarly activity through the SAJS

## AUTHOR:

Roseanne Diab<sup>1</sup>

## AFFILIATION:

<sup>1</sup>Academy of Science of South Africa, Pretoria, South Africa

## CORRESPONDENCE TO:

Roseanne Diab

## EMAIL:

roseanne@assaf.org.za

## POSTAL ADDRESS:

Academy of Science of South Africa, PO Box 72135, Lynnwood Ridge 0040, South Africa

## KEYWORDS:

open access; digital publishing; multidisciplinary journal; scholarly publishing; South Africa

## HOW TO CITE:

Diab R. ASSAf: Promoting scholarly activity through the SAJS. *S Afr J Sci.* 2016;112(1/2), Art. #a0141, 2 pages. <http://dx.doi.org/10.17159/sajs.2016/a0141>

The *South African Journal of Science (SAJS)* is an integral part of the Academy of Science of South Africa's (ASSAf's) core activities, responding directly to one of ASSAf's five strategic goals, namely the promotion of innovation and scholarly activity in South Africa, with a special emphasis on all forms of interdisciplinarity based on the core and common role of empirical enquiry. ASSAf's role as publisher of the journal is critical to the strategic direction and successful implementation of its scholarly publishing and open access activities. The SAJS is at the forefront of many new initiatives that ASSAf is introducing and that will influence the future of scholarly publishing in our country.

Dating back to 1903, the SAJS was first published as the proceedings of the annual meetings of the South African Association for the Advancement of Science, now known as the S<sub>2</sub>A<sub>3</sub>. A glance at the contents page of the first volume reveals a focus on topics not much different from the burning issues of today; there was a presidential address by Sir David Gill, a Scottish astronomer who spent much of his career in South Africa, and articles titled 'The Moral Education of Children in Schools', 'The Life of the City', 'Sewage Disposal in the Cape Colony', and 'Some Aspects of South African Forestry'.

In August 1947, the annual report became a monthly publication and the journal was published under the auspices of various bodies, including Macmillan Journals in London (the then publisher of *Nature*), the Associated Scientific and Technical Societies of South Africa, the Foundation for Education, Science and Technology, and the National Research Foundation. There were some difficult times over the years, but the journal sustained itself without interruption.

In 2002, a nascent ASSAf took the bold step of assuming responsibility for the publication of the SAJS, with the aim of building the journal's reputation as an independent scholarly journal of the multidisciplinary type. At the time the journal was in sound editorial hands, but needed a visionary approach to take it from a paper-based journal that had a relatively narrow focus of 'natural scientific' interest to one that is entirely digital and open to good papers from all empirical fields of enquiry. In keeping with the dual content of the world's leading multidisciplinary journals, the SAJS now aims to be the foremost repository of editorial comment, scholarly debate and review, and science and technology policy analysis relevant to South Africa. This lofty goal is still a 'work in progress', but significant strides have been made.

The journal's success is due in no small measure to the dedication of its succession of recent editors. The first of these was Dr Graham Baker, who arrived in South Africa in 1972 after a science publishing career with *Nature* in London. He set about the demanding task of taking the journal from a barely viable condition to a flagship multidisciplinary journal modelled on *Nature*. For 36 years, as full-time Editor, he dedicated himself to the establishment of the SAJS as a high-quality, internationally significant journal that showcased South African natural science research to a global audience.

Towards the end of 2008, ASSAf introduced a new editorial model, with Prof. Michael Cherry as part-time Editor-in-Chief, supported by ten part-time Associate Editors in various disciplines. It was at this time that the focus of the SAJS was also deliberately broadened, specifically targeting the hitherto neglected humanities and social sciences. Since then, submissions in the fields of the humanities and social sciences have steadily increased, warranting a recent decision to expand from one portfolio into two (each managed by an Associate Editor).

In 2009, the SAJS fully embraced open-access publishing and was the first South African journal to be uploaded onto ASSAf's new open-access platform, SciELO SA. Changes also took place in publishing mode, with the adoption of an online manuscript management system for the submission, peer review and publication of papers, now outsourced to OpenJournals Publishing. Digital publishing introduced new file formats (HTML/XML and EPUB), reference linking and DOIs for enhanced searchability and user friendliness. Embarking on digital publishing was a steep learning curve but was achieved through the dedicated efforts of a large number of people who set the SAJS on a course from which there was no turning back.

Dr John Butler-Adam took over as part-time Editor-in-Chief in November 2012. His assumption of the editorship coincided with ASSAf once more itself taking control of the publication of the journal – a move which brought with it exciting possibilities, but also a new set of challenges. At the same time, there was a rapid increase in the number of articles being submitted for publication from all parts of the world as the accessibility and reputation of the journal began to increase. The increased responsibilities in the ASSAf secretariat, together with the opportunities and demands of digital publishing, saw the creation of the post of Online Publishing Administrator in 2013, with Ms Nadine Wubbeling appointed part-time and later full-time in the role.

Digital publication of the SAJS has steadily become the dominant mode. Initially, the SAJS was published in dual mode, both print and digital, with the printed journal distributed free to ASSAf Members. In 2014, a decision was taken to discontinue the free distribution of hard copies and to focus on electronic distribution. Effectively, the hard copies were reduced from 700 to just above 100. Recently, the ASSAf Council approved the discontinuation of the print version in favour of electronic distribution via the bimonthly 'Highlights of the latest issue' emails that now reach over 9000 recipients. These emails include links to the full digital issue in three formats (PDF, EPUB and flip ebook) and are sent at a minimal cost of ZAR0.14 per recipient.

Digital publication has also ushered in new and sophisticated possibilities in terms of impact monitoring, so important in this era of ensuring effective utilisation of state funding and competing demands for resources. Using Google Analytics, one can capture data on readership by country, city and article. The *SAJS* has a developing social media presence: it has a Facebook page with 831 'likes' as of December 2015 and an active Twitter account, with 553 followers and 730 tweets. Media reporting of *SAJS* articles is monitored closely and is showing a pleasing growth.

It is planned to introduce Altmetric software that will also include alternative metric 'counts' and statistics from social media, blogs, news coverage and other online sites. Article-level metrics such as full-text downloads and citations will also be included for each new article.

The all-important Thomson Reuters Web of Science impact factor has shown a steady rise and is standing at 0.96 in 2015, up from 0.5 in 2010. The *SAJS* at 1848 cites ranks second among the South African journals on Thomson Reuters Web of Science in terms of citations.

The face of the *SAJS* has thus changed dramatically over the past few years. Much has been learned, and ASSAf is now able to assist other South African scholarly journals to benefit from the enormous advantages of open access and digital publishing through the National Scholarly Editors' Forum.

While the *SAJS* is currently in a very healthy state and is a publication of which the Academy is justifiably proud, there are still challenges to address, of which two can be regarded as foremost.

The first is the need to encourage submissions of higher-quality papers. As submissions have increased, it has become possible to be more selective, thereby increasing the rejection rate, which is often used (in a rather perverse way) as a measure of the quality of a journal. The

formal rejection rate, excluding those which are informally submitted to the Editor-in-Chief for an opinion, is currently hovering just above 85%, which is a healthy sign. Submissions have increased at an unprecedented rate, from just above 200 in 2009 to almost 500 in 2015. At this annual growth we are expecting to reach 600 submissions in 2017. While this is a measure of success and improved profile, it comes with numerous practical problems relating to costs, capacity constraints, publishing backlogs, reviewer fatigue etc., all of which have to be addressed.

A second and related challenge is the need to position the *SAJS* as a 'destination journal of choice'. All too often one hears the refrain that international specialist journals should be the first choice, with the *SAJS* placed fairly low in terms of preferred journal. The challenge is one of influencing the mindset of South African researchers such that the *SAJS* is not viewed as a last-resort local journal to which one can turn after one's paper has been rejected by a specialist international journal, but as an outlet where an author deliberately crafts an important paper for a multidisciplinary audience. It should be seen as an opportunity for scientists to communicate their work beyond a specialist audience and to make their work broadly accessible to a multidisciplinary audience and thereby improve impact on society. Increasingly, scientists need to pay attention to effective and broad-based communication of their results; it is no longer considered sufficient to confine results to specialist journals. As the pressures to communicate in the popular press and through social media are becoming more and more important, so the multidisciplinary journal is firmly establishing its niche. The role of South Africa's foremost multidisciplinary journal is fundamental to building the reputation of South African science and packaging it in a way that is positioned midway in the spectrum that extends from specialist to popular. There is still a journey ahead to realise this ambition, but the *SAJS* is now on an upward trajectory that could see this being achieved sooner rather than later.



# Breast cancer: When do you stop reading the literature?

## AUTHOR:

Brenda Wingfield<sup>1</sup>

## AFFILIATION:

<sup>1</sup>Faculty of Natural and Agricultural Sciences, University of Pretoria, Pretoria, South Africa

## CORRESPONDENCE TO:

Brenda Wingfield

## EMAIL:

Brenda.Wingfield@up.ac.za

## POSTAL ADDRESS:

Faculty of Natural and Agricultural Sciences, University of Pretoria, Private Bag X20, Hatfield 0028, South Africa

## KEYWORDS:

mammogram; cancer research; oncology; scientific literature

## HOW TO CITE:

Wingfield B. Breast cancer: When do you stop reading the literature? *S Afr J Sci.* 2016;112(1/2), Art. #a0135, 3 pages. <http://dx.doi.org/10.17159/sajs.2016/a0135>

Science and research consume a substantial – some might say the greater part – of my life. It is an all-consuming passion that many fail to understand. Professionally, I divide my time between managing a substantial research programme and running what is arguably the largest science faculty in South Africa, if not the continent. I am not only passionate about my research but also about sharing this passion with others, particularly young researchers, and I believe I am good at what I do.

A year ago, and perhaps reluctantly given many other responsibilities, I arranged for one of my least enjoyable responsibilities – an annual mammogram. This year my visit took a bit longer than usual because the radiologist found a mass ‘he did not like’ and that required a biopsy. At this point I did not give the situation much further thought. A week later I received a phone call to inform me that I had cancer. This phone call was the start of a very frustrating period in my life; perhaps ironically made especially difficult for me because of my research training. It heralded a time at which I entered a world both foreign and frightening – that of being a cancer patient.

As a scientist, my life follows a very logical route and one that is largely based on evidence. I have been trained to ask strong questions and to test hypotheses that might lead to answers; this process is the central core of my professional life and I debate everything, even with myself. This philosophy inevitably spills over into my personal life. My first approach to any problem is to read and understand the literature and to then make decisions based on available evidence. I am sure that this approach resonates with most scientists; it is after all what we say in our conversations with our graduate students: ‘One cannot embark on any project until one has read the relevant literature.’ I thus entered the world of being a cancer patient in the same manner: I started reading the appropriate literature.

I read first to understand my diagnosis and then to try to determine what might be a reasonable treatment. Initially what I read seemed completely logical; but the more I read the less logical it all became. The scientific literature, while encompassing many details that serve to explain some of the key issues, is also crammed with statistics. And therein I encountered one of my greatest frustrations: it was not possible to know which statistic I represented. Was I one of the very small percentage that is unlucky and would have a recurrence of the cancer? Or one of those representing a much larger percentage who will live a long life in the absence of such a recurrence?

As I delved further into the literature concerning breast cancer diagnoses and treatments, I was reminded rapidly that statistics are useful when one is conducting research but they are not particularly helpful when you yourself are one of those statistics. As a scientist I rely on probability testing for many decisions. Thus it was incredibly difficult for me to encounter a part of my life in which statistics were not particularly helpful. In terms of reaching decisions based on clearly understanding my diagnosis, I found that my training in science had failed me. It may well be that my science-based world is inordinately black and white.

Reading the literature on breast cancer and its treatment highlighted for me the fact that in many respects these aspects represent a debate – one that is clearly ongoing, which is perhaps not surprising when I think of my own research exploits. But it quickly became evident to me that for many of my questions relating to breast cancer there are no clear answers. I realised that it is one thing to do research oneself and to be part of the debate, in which someone will eventually reach an acceptable answer, but another thing to be stuck in a moment in time when one desires a clear answer immediately but learns from the literature that there is no consensus. It is a situation that is nothing less than confusing and frustrating.

As a scientist, one’s training is to read the literature and then, at that edge of the unknown, to design experiments to answer key questions. Essentially being part of what must be a larger ongoing experiment is a very uneasy place in which to find oneself. I realised that, concerning my own diagnosis, it would be much more comfortable to know very little, to blindly believe the views of the medical profession and to simply accept whatever advice they were able to provide. Subsequent to my diagnosis, I have engaged in discussions with other scientists who have cancer. Some simply have stopped reading the literature because they felt that it only added to what is already an emotionally and physically difficult situation. Interestingly, some of my colleagues have given me exactly this advice: ‘accept and hope’. To do so would be contrary to my background and nature and I have not been able to follow such a route; in fact, I have found myself reading increasingly more regarding breast cancer and the treatment options.

My family, a supportive scientist husband and two great children, have told me that they perceive me as having embarked on a second PhD – on breast cancer and its treatment. I have trained 40 PhD students during my research career and I have a clear understanding of how one should best pursue the process of interrogating the literature for a PhD: read and understand what is known, and only then is it possible to determine the gaps in knowledge to then formulate questions that can be answered. In terms of my own diagnosis, this approach has been patently without value. I have to admit that I have at times thought seriously about whether I should not be considering undertaking some research on my own tumour. I certainly have the professional skills to do a number of very interesting experiments. For example, it would be possible to determine whether the tumour was over- or underexpressing a range of genes. I am also very conscious of the fact that some of the experiments that I would consider doing will probably become standard practice within the next 5 years. Again, because of my profession as a molecular geneticist, I know more than the average patient. In contrast to those patients, this knowledge means



that I cannot fail to become preoccupied with the many possibilities that would make the decisions that I need to make regarding long-term diagnosis so much easier.

The frustrations that I have experienced regarding my diagnosis of breast cancer and its treatment are not limited to cell biologists and those people that understand genetics. In the waiting room of the radiation oncology practice where I received treatment, I entered into a conversation with a geologist, one who I happen to know professionally. His preoccupation with the treatment he was receiving was more in regard to the impact of the radiation itself, rather than the genes that I was considering. He suggested that the treatment of cancer using radiation might be thought of as a blunderbuss approach to the problem. I thus have come to believe that scientists who are faced with a diagnosis of cancer are likely to become preoccupied with that aspect of their diagnosis and treatment that is most closely linked to their own area of expertise, which makes me think of the saying 'give a man a hammer and he will find a nail'. Central to our training as scientists is the search for answers, and, in the case of cancer but perhaps all of biology, this search clearly remains a work in progress. My difficulty (which I am sure is true of other scientists too) is having to accept this fact outside of my own laboratory and concerning my own body.

As a scientist and researcher, I have learned to debate and discuss all possible options based on available evidence. When this evidence does not exist, one then proceeds to design experiments to provide the needed evidence. This is clearly not how medicine is practised. At some level, I feel some sympathy for the medical professionals who have had to deal with me. I must guess that most do not have the difficult experience of encountering a patient bearing files packed with the latest literature. In some way, even medical specialists are 'body mechanics' that must perform to some predetermined standard operating procedures. These procedures probably change over time and as new discoveries are made, but they seldom can be sufficiently up to date with the most recent scientific literature. One of my challenges has thus been to decide whether I should challenge the medical professionals treating me when I know that some of the procedures recommended are not necessarily up to date. I suppose that it is unrealistic to expect that the average specialist is reading the latest literature and applying it to their clinical practice. And I am sure that those specialists in private practice do not have access to the scope of literature that I am used to and expect to interrogate in my professional life.

As I have dealt with my breast cancer, I have found myself in a rather uncomfortable space between the reality of my condition, which needs to be dealt with, and my need to understand exactly what is wrong with me in scientific terms. I have discovered that the gap between what the medical profession understands and what I would like to understand is rather wide. This statement does not reflect negatively on the medical profession but rather highlights the fact that, despite huge leaps in knowledge, there is a vast amount regarding cancer that we do not fully understand, and certainly not at the level at which I desire understanding.

This period of my life dealing with breast cancer has been focused on making decisions about treatment, but it has also been a period in which I have *painfully* realised that it is not possible to make these decisions based on evidence. The first decision I had to make was whether I should have the 'mass', a potential tumour, removed. The medical opinion on this was clear: 'cut it out'. As I delved increasingly deeper into the literature, I became increasingly less convinced that this suggestion was in itself as clear as I was led to believe. The biopsy indicated that my cancer was slow growing, and if I had not had the mammogram I likely would not have noticed it at all; there was no lump, there were no symptoms and there was no other indication of a problem. In fact, based on the literature that I have consumed, I am not sure that my decision to have the mass removed was actually the correct one. The medical advice that I had received was overwhelmingly in favour of removal and I followed that advice. Not unimportantly, I was also financially able to afford the operation, and other than the obvious inconvenience and pain that accompanies surgery, undergoing the procedure was not too difficult.

After undergoing the operation and receiving a diagnosis on the tumour itself, the situation became a bit clearer. Yet the way forward seemed even less obvious than it had before. My tumour was very slow growing, as had been previously diagnosed. The entire tumour had been removed and the lymph glands that had also been removed were clear of cancer. This was good news, but it also meant that the decisions associated with the way forward were not so obvious, at least to me. It was at this time that I embarked upon an even more serious reading of the scientific literature. With my background in genetics and biochemistry, I could easily understand the medical literature despite the fact that I do not usually read it; my reading largely being focused on the genetics of fungal pathogens of trees.

I understand genetics and because I work in a genetics department I have a general idea of the latest breakthroughs in human genetics. In terms of breast cancer, the *brca1* and *brca2* genes have featured significantly. I was aware of the fact that these genes account for a very low percentage of the breast cancer cases that are diagnosed. I also know – because I have exposed my own genome to analysis using the million SNP markers – that I do not have either of these variants. What was rather alarming was that none of the medical specialists I encountered was at all interested in my knowledge of genetics or what I knew about my own genome. The surgeon who operated on me was more interested in cutting out the tumour and the oncologist was more interested in treating me post-surgery than discussing information from my own genome analysis. Not surprisingly, this lack of interest did not make my decisions any easier.

I have found in my recent interaction with the medical profession, and I add here that I have close friends in this field, that there is a requirement that patients sign forms at every turn, confirming that they consent to the various procedures that will be followed. I can understand that this requirement is important from an ethical point of view and I support the approach. What I am not so sure about is the concept of 'informed consent'. I certainly had a rather circular discussion with one of the medical professionals treating me before he finally stated that he needed to have my informed consent. My reaction was that he had my consent, but while he had 'informed' me as far as he could, I did not believe that I, and less likely other patients, understood enough to be making a decision based on 'informed consent'.

It might be argued that we live in an age in which we have an inordinate amount of information available to us (even whilst convalescing). I thus found myself in a rather circular manner, reading the scientific literature and then resorting to the 'lay' literature. Neither were that satisfying; the former because I could not derive the answers I needed and the latter because there were too many generalisations. My 'pet peeve' has become the common statement 'radiation kills cancer cells'. Well actually radiation kills all cells, and most often cells that are dividing are more sensitive to radiation. It is true that cancer cells on average divide more rapidly than many (but not all) other cells in the body. These generalisations, which also are very misleading, irritated me. I understand that their purpose is to simplify matters and to enable easier understanding for the general public. But I am not the general public and I know too much to feel appeased. This irritation sent me back to the scientific literature, which still contained inordinately few answers.

The decisions that I needed to make post-surgery – clearly underpinned by 'informed consent' – were to agree to radiation therapy and, because my tumour was oestrogen receptor positive, to consider taking an oestrogen blocker. I have already mentioned my sentiments regarding radiation – in my opinion, using the words therapy and radiation in the same sentence is oxymoronic. But I needed to overcome this viewpoint and to engage in a sensible discussion on the topic. The current literature, while extensive, seems only to cover the 5–10-year horizon. I would like to live longer than this time frame. In order to make this decision, I spent a considerable time reading, discussing and eventually debating the issue with a pleasingly accommodating oncologist. The average cancer patient will do exactly what their doctor suggests. The average scientist needs evidence, which in this case was not really available. I took a very long time to make my decision and in the end actually took a rather cynical approach. What I did was to leave the decision to my



medical aid. I reasoned that if the medical aid was prepared to pay for the treatment I would accept it. I realise that this decision was based on a flawed argument; the decisions of medical aid schemes are based on statistics and one cannot really know which data point one represents in this process.

My journey with cancer has only just started. I sincerely do not wish to repeat the last year of my life. I also am hoping that the next few years

will not involve making quite as many impossible decisions as those that I recently have had to face. All indications (and comfortingly the view of a friend who is also a physician and medical scientist) are that I am unlikely to die from my cancer. Of course one doesn't actually know this for sure until one has died – from another cause. I suspect that I will continue to read the scientific literature on breast cancer in addition to the other scientific literature that I read. I will not be submitting my 'second PhD' for examination, partly because it will never be complete.



# Urban farming as a possible source of trace metals in human diets

## AUTHORS:

Joshua O. Olowoyo<sup>1</sup>  
Gladness N. Lion<sup>1</sup>

## AFFILIATION:

<sup>1</sup>Department of Biology, Sefako Makgatho Health Sciences University, Pretoria, South Africa

## CORRESPONDENCE TO:

Joshua Olowoyo

## EMAIL:

woleolowoyo@yahoo.com

## POSTAL ADDRESS:

Department of Biology,  
Sefako Makgatho Health  
Sciences University, PO Box 60,  
Medunsa 0204, South Africa

## DATES:

**Received:** 08 Dec. 2014

**Revised:** 08 Apr. 2015

**Accepted:** 16 June 2015

## KEYWORDS:

pollution; vegetables; soil;  
human health

## HOW TO CITE:

Olowoyo JO, Lion GN. Urban farming as a possible source of trace metals in human diets. *S Afr J Sci.* 2016;112(1/2), Art. #2014-0444, 6 pages. <http://dx.doi.org/10.17159/sajs.2016/20140444>

Rapid industrialisation and urbanisation have greatly increased the concentrations of trace metals as pollutants in the urban environment. These pollutants (trace metals) are more likely to have an adverse effect on peri-urban agriculture which is now becoming a permanent feature of the landscape of many urban cities in the world. This review reports on the concentrations of trace metals in crops, including leafy vegetables harvested from different urban areas, thus highlighting the presence of trace metals in leafy vegetables. Various pathways of uptake of trace metals by leafy vegetables, such as the foliar and roots, and possible health risks associated with urban farming are discussed and various morphological and physiological impacts of trace metals in leafy vegetables are described. Defensive mechanisms and positive aspects of trace metals in plants are also highlighted.

## Introduction

Urban farming can be defined as the act of cultivating food crops, mostly vegetables, wherever land is available around or in the immediate vicinity of a major city.<sup>1,2</sup> Farming activities around urban areas have become a prominent feature of some urban city landscapes, especially in developing countries.<sup>3</sup> There are several reasons for farming activities around peri-urban areas and these include easy access to markets and transportation of goods. In some countries, poor urban households use urban farming to increase their household income by selling the yield or surplus to reduce part of daily expenses.<sup>4</sup> The other probable reason for the increase in urban farming is the shortage of land. This involves farming areas previously used for either household or industrial waste,<sup>5</sup> supported by the notion that the soil will be very fertile. Urban farming also supports the campaign for organic farming. Consumers, in most cases, evaluate quality of leafy vegetables on their dark green colour and on size of the leaves as opposed to where farming activities have taken place.<sup>6</sup> However, in urban city centres where farming activities are carried out around industrial areas, safety around the consumption of these vegetables cannot be guaranteed because various disposal practices often cause the accumulation of potentially toxic trace elements in the soil.<sup>7,8</sup>

Anthropogenic activities such as mining, emissions from vehicles, wrong agricultural practices and improper waste disposal are major sources of trace metal pollution in the urban environment.<sup>5,9</sup> It was recently discovered that a large percentage of toxic trace metals find their way into the human diet through consumption of vegetables and agricultural products.<sup>5-14</sup>

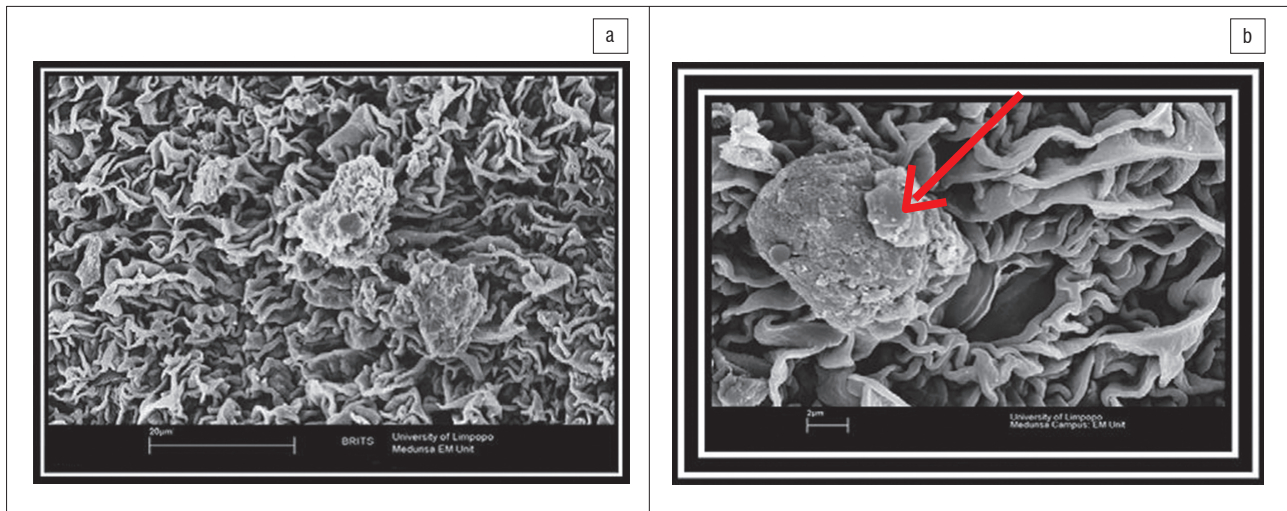
In acceptable limits, trace metals play an important role in the health and physiological activities of plants, animals and humans.<sup>15</sup> They are required in minute quantities as natural components of the environment. For example, zinc is an essential element required in minute quantities in living organisms, but when supplied in high quantities, it can be toxic to plants, producing purplish-red coloured leaves which is a symptom associated with phosphorus deficiency. Zinc may also cause chlorosis in younger leaves which may extend to older leaves.<sup>16</sup> In humans, excess zinc may lead to metal poisoning and growth retardation. Excess nickel and lead may result in increased production of reactive oxygen species and membrane permeability disruption in plants. In all, concentrations of heavy metals above the required limits in plants are known to cause various deleterious effects on several plants systems such as the photosynthetic ability of the plants, mineral uptake and interactions with the water regime from the soil. In humans, excess lead may affect the functions of the liver and kidneys.<sup>10,17,18</sup>

The present review will establish, amongst others, the various pathways by which trace metals may be taken up by plants, the effects of trace metals on the morphology and physiology of the plants, the positive aspects of trace metals in plants and also the possible health risk for humans and livestock if trace metals are ingested in high concentrations.

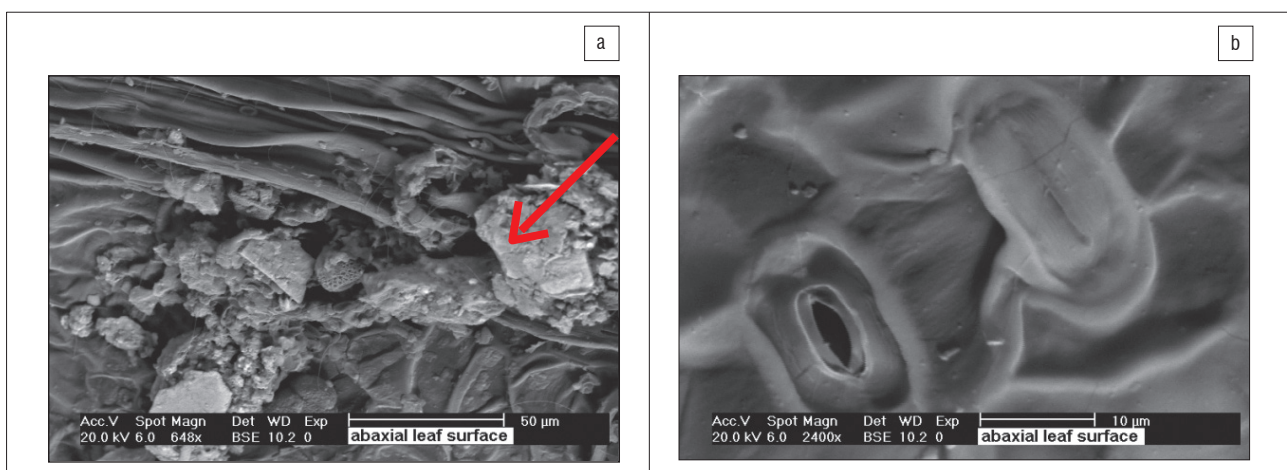
## Evidence of trace metals in plants

High deposition and accumulation of trace metals in the edible part of root and leafy crops has been reported in the literature.<sup>19,20</sup> Vegetables are capable of accumulating trace metals from polluted soil and also from surface deposition onto their shoots in polluted atmospheric environments.<sup>14</sup> Trace metals in the air have been reported to significantly influence total metal concentration of vegetable plants, especially when washing is not thoroughly done.<sup>18</sup>

Atmospheric fallout is one of the chief contributors to heavy metal uptake by plants through the stomata. The stomata openings are located on the surface of the plant leaves and perform multiple functions that include water regulation in the plant. Particulate matter from atmospheric fallout may be found deposited on the leaf surface and find a way into the leaves through the stomata. Smaller particles from atmospheric fallout may be incorporated into the leaves, whereas large agglomerates are trapped on the surface wax.<sup>21,22</sup> The extent of uptake and the pathways involved may depend on the plant species and on the metal involved.<sup>23</sup> The opening and closing of stomata may provide entrance for trace metals, blocking the stomata in some cases and may ultimately lead to the death of the plant.<sup>23</sup> Entrance of metals via the foliage parts of plants was noted to be one of the major pathways by which metals enter leaves in a polluted area.<sup>24</sup> Figures 1 and 2 indicate the presence of trace metals around the stomata.



**Figure 1:** Presence of pollutants around the stomata of leaves collected for a study conducted in Pretoria, South Africa at (a) 2000X magnification and (b) 3000X magnification.<sup>9</sup> The stomata are the bean-shaped structures and the arrow is pointing to the pollutants around them.



**Figure 2:** Presence of pollutants around the stomata of leaves collected for a study conducted in Belgrade, Serbia.<sup>25</sup> at (a) 648X magnification and (b) at 2400X magnification. The stomata are the bean-shaped structures and the arrow is pointing to pollutants around them.

In a separate study on lead uptake by lettuce leaves, it was discovered that particulate matter deposited on plant leaves may be retained by cuticular waxes and trichomes, while some of the metals contained in particulate matter can penetrate inside plant tissues.<sup>26</sup> Micro-X-ray fluorescence, scanning electron microscopy coupled with energy dispersive X-ray microanalysis, and time-of-flight secondary ion mass spectrometry were used to investigate the localisation and the speciation of lead in the leaves of different plants around a copper smelter company and it was found out that lead-enriched particulate matter was present on the surface of plant leaves.<sup>24</sup> The study further reported on biogeochemical transformations on the leaf surfaces with the formation of lead secondary species ( $\text{PbCO}_3$  and organic lead).

### Toxicity of trace metals in plants

The effect of trace metals on the epidermis was demonstrated on the young leaves of soyabean.<sup>26,27</sup> The young leaves did not show visual symptoms when exposed to cadmium ions and the authors observed that cadmium ions did not have any effect on the closing and opening of the stomata. However, findings have revealed that interactions of cadmium ions with  $\text{K}^+$ ,  $\text{Ca}^{2+}$  and abscisic acid showed strong interference with the guard cells.<sup>28,29</sup> The application of cadmium on the leaf surfaces influenced the number of stomata, decreasing the number of stomata and stomatal openings on the leaf surface of the young leaves of soybeans.<sup>27</sup>

Several authors have also reported on the adverse effect of lead in plants. Lead toxicity is said to have a negative effect on biomass production in plants as it affects chlorophyll biosynthesis and photosynthesis.<sup>30,31</sup> Increased concentrations of lead may also inhibit or delay enzyme activity changes in membrane permeability and water disturbance in plants, which may affect the growth of the plant negatively.<sup>32</sup>

Cadmium may penetrate the root via cortical tissue and reach the xylem through either the apoplastic or symplastic pathways possibly resulting in cadmium toxicity in plants.<sup>33,34</sup> Soil pH is known to influence cadmium uptake and transportation. Uptake of cadmium by corn was lower in acid soils with high organic matter content. Cadmium has also been found co-accumulated with zinc in the aerial parts of *Arabidopsis halleri*.<sup>35</sup> Toxicity of nickel affected the seedlings of *Pisum sativum* by changing the potassium uptake and water content,<sup>36</sup> showing that higher concentrations of nickel may lead to a reduction in plant growth, leading to oxidative stress.

When leafy vegetables are properly washed, the concentrations of chromium and lead may be reduced. However, the exogenous contamination of leaves may not be reduced in some instances owing to the nature of the vegetables, as was in the case with *Gynandropsis gynandra* L. that showed a marked tendency to accumulate lead and chromium.<sup>5</sup> A similar observation was noted in a study by Gabrielli and Sanità di Toppi<sup>37</sup> where the dominant pathway for most trace elements

to vegetable roots was from the soil, while trace elements in vegetable leaves appeared to originate mostly from the atmosphere. The result further indicated that high accumulation of trace metals such as lead, cadmium and chromium were the result of atmospheric deposition.

In all, it was evident that trace metals could reach the edible parts of vegetables, especially the leafy parts, through atmospheric deposition and could also be translocated to various parts of the plants via the root system.

### Trace metals uptake mechanism by plants

The process of uptake, translocation and bioaccumulation of trace metals in plants could be influenced by a number of factors such as climate, atmospheric deposition, the concentration of trace metals in soil, the nature of the soil in which plants are grown (soil pH, soil organic matter content and soil texture), the degree of maturity of the plants at the time of harvest and the type of trace metals.<sup>15,37-39</sup> Among crop types, leafy vegetables such as lettuce and cabbage have the greatest ability to take up trace elements from the soil.<sup>40</sup> The mobility of trace metals in soil is favoured mostly under acidic conditions. Treating soil with lime in order to reduce soil acidity reduces the bioavailability of trace metals.<sup>41,42</sup> Therefore, low pH levels are effective in the remediation of polluted soil, while the soil organic matter ensures the availability of trace metals to the plant.<sup>42</sup> On the other hand, at high levels of soil pH, the formation of soluble organometallic complexes may increase metal solubility although this is not true for calcareous soil.<sup>43</sup>

An increase in the amount of soil organic matter helps plants minimise trace metal absorption.<sup>44</sup> The introduction of organic matter amendments in conjunction with lime had been used to assist in immobilising trace metals.<sup>45</sup> However, the effects of organic matter on the bioavailability of trace metals in soil depend on the nature of the organic matter, microbial degradability, salt content, soil type and the particular heavy metal.<sup>46</sup> In most cases, metallic elements are actively retained by lands that are rich in organic matter. The retention of trace metals in the soil affects the mobility of these elements and interferes with the uptake process. The bioavailability of trace metals is much lower in the presence of manure when compared to humified compost, suggesting that different types of organic matter may affect mobility of trace metals differently.<sup>47</sup> This may be as a result of the ability of the matter used to redistribute trace metals from soluble and exchangeable forms to fractions associated with organic matter or carbonates and the residual fraction.<sup>48</sup>

In the root area where the plant interacts with the soil, metal ions cannot move freely across the cellular membranes. The membrane structure of the plant root is lipophilic, which requires that the transport be facilitated by proteins with transport functions.<sup>49,50</sup> In the rhizosphere, metal transport is carried out by two processes known as bulk flow and diffusion.<sup>51,52</sup> This process of uptake by plants is soon followed by a process controlled by root pressure and leaf evaporation, called transpiration. Through transpiration, the plant is able to absorb the trace metals through the roots via the xylem to the shoot of the plant. The process is dependent on water demand by the leaves in the aerial part of the plant.<sup>49</sup>

Uptake of trace metals by higher or vascular plants is often through the root system, but can also occur through the leaves. It may therefore be difficult to distinguish whether the metals found in the plant tissues were originally from the air or soil.<sup>53</sup> Plants that are tolerant to high levels of trace metals (also called hyperaccumulators) have the capacity to remove contaminants from the soil. One example of such a plant is *Thlaspi caerulescens* that has been used for phytoremediation of soils, especially in areas previously polluted by mining activities.<sup>41</sup> The mechanism for hyperaccumulation remains unclear though it is generally believed to involve three major phases involving rapid uptake of metals by the roots, high rate of translocation from roots to shoots and high storage capacity by vascular compartmentalisation.<sup>41</sup>

Metal uptake by plants is affected by metal solubility and availability in the soil. In a situation where the level of trace metals in soil is very high, the release of root exudates and acidification are common mechanisms that are used by plants to modify the root area to acquire nutrients from

the soil.<sup>41,49</sup> In the case of nutrient movement across the biological membrane, plants have developed a specific mechanism mediated by proteins, for uptake, translocation and storage of the nutrients.<sup>42,49</sup> Membrane transporters are equipped with a structure that binds ions before transportation. This structure is receptive only to certain ions and as such is specific in their mode of action.<sup>49,54</sup> The transmembrane structure then facilitates the transfer of bound ions from extracellular space through the hydrophobic environment of the membrane into the cell. Despite the presence of this structure, only a fraction of the total amount of ions associated with the roots are finally absorbed into plant cells.<sup>54</sup> The other form of metal uptake in plants, apart from binding to the cell wall, is sequestration in cellular structures such as the vacuole, though this may make the heavy metals unavailable for translocation to the shoot.<sup>55</sup>

The evaporation of water from the leaves may also affect plant uptake and accumulation of trace metals. The evaporation process serves as a pump for more nutrients and other substances to be absorbed into plant roots. This process, called evapotranspiration, moves water and contaminants into the plants.<sup>42</sup> The accumulation of metal contaminants is mostly assisted by microorganisms, fungi and bacteria that live in the root area. These microorganisms in the rhizosphere and closely associated plants may contribute to the mobility of metal ions. At the same time, plant roots release nutrients that sustain a rich microbial community in the root area, thus establishing an important symbiotic relationship between soil microorganisms and plants.<sup>42,49</sup> In order to facilitate the transport process, several families of proteins are involved namely (1) influx transporter families such as zinc, a regulated transporter, iron, a regulated transporter protein, yellow-stripe and natural resistance associated macrophage protein and (2) efflux protein families such as cation exchanger, ATB-binding cassette and cation diffusion facilitator.<sup>42,56</sup> Because these proteins are substrate specific, the comparison between influx and efflux transporters revealed that efflux proteins export metals from the cytoplasm while influx proteins take up proteins from the soil or medium.<sup>56</sup>

Depending on their ability to adapt and reproduce in soils heavily contaminated with trace metals, higher plant species can be divided into two main groups. The two groups are the pseudometallophytes (plants that grow on both contaminated and non-contaminated soil) and absolute metallophytes (plants that grow only on metal contaminated and naturally metal-rich soils).<sup>57</sup> The use of *Raphanus sativus* (a pseudometallophyte) for example, demonstrated the potential for root uptake in lead contaminated soil.<sup>58</sup> Baker<sup>58</sup> showed that radish is a hyperaccumulator plant that can concentrate trace metals in different plant parts. It was also demonstrated that radishes are effective for remediation of polluted soil through their potential to extract metals from soil up to a certain level of concentration.<sup>58</sup> The ability of plants to accumulate metals, thereby remediating metals, is directly proportional to the presence or availability of microorganisms in that plant's rhizosphere.<sup>59</sup> In the study it was explained that microbial communities such as fungi, bacteria and other microbes are capable of altering the soil environment and as a result will translocate, absorb or sequester contaminants such as trace metals.<sup>59,60</sup>

Over the years, more than 400 plant species with the ability to take up high levels of heavy metals in soil and water have been identified. *Thlaspi* spp., *Brassica* spp., *Sedum affredii* and *Arabidopsis* spp., among others, were studied.<sup>61,62</sup> The use of vegetable plants has also been demonstrated by some researchers.<sup>63</sup> For example, *Amaranthus dubius*, also known as *morogo* or wild spinach in South Africa, was found to have the ability to take up and translocate metals such as chromium, mercury, arsenic, lead, copper and nickel to the aerial parts of the plant.<sup>64</sup> Some medicinal plants such as *Datura stramonium* and *Amaranthus spinosus* are capable of accumulating some trace metals in their tissues.<sup>60</sup>

It is believed that trace metals can help plants protect themselves from diseases and biological stress.<sup>65,66</sup> If a metal becomes more toxic to a pathogen than to the plant, the metal can hamper the virulence of the pathogen and can increase the resistance of the plant to the biotic stress<sup>65</sup> by suffocating the pathogen in the plant. The excess trace



metals found in the plant after the pathogen has been suffocated will then be redirected to normal growth.

The production of high levels of reactive oxygen species can adversely affect the plant. Therefore, plants have developed a defensive mechanism that involves glutathione in the detoxification of reactive oxygen species through the ascorbate–glutathione cycle.<sup>67</sup> During exposure to high levels of trace metals, accumulated metal ions are detoxified by phytochelatin that are produced from glutathione in the plant. These metal ions are then bound to phytochelatin to form complex structures that are sequestered or compartmentalised in the vacuole.<sup>49,67</sup>

Exposure to trace elements such as mercury, cadmium, lead and nickel in the soil, encourages the plants to formulate steps to counteract the effects of these toxins. Defensive mechanisms largely prevent the metals from getting inside the cells, but for metals that find a way into plants cells, they are neutralised and sequestered.<sup>68</sup>

## Quantifying human risk associated with trace metals in plants

In a bid to quantify the likely health hazard associated with vegetables that are high in concentrations of trace metals, the target hazard quotient method (THQ) was developed and has been used by several authors.<sup>69,70-72</sup> Human health risks associated with these metals can be assessed based on the THQ method,<sup>41</sup> which takes into account the concentration of trace metals in food, the frequency of exposure, and the individual's age, body weight and frequency of consumption of the contaminated food. Should the THQ calculated for both adults and children exceed 1 (THQ > 1) then a potential risk to the consumer will be suspected.<sup>5,14,41</sup> Mercury was recorded as the major health risk contributor in children and chromium as the least contributor. The method for calculating the THQ is:

$$HQM = ADDM / RfDM, \quad \text{Equation 1}$$

where  $ADDM = (DI \times MFveg) / WB$ .<sup>14</sup>

ADDM is the average daily dose (mg.kg/day) of the metal and RfDM is the reference dose (mg.kg/day). RfDM is defined as the maximum tolerable daily intake of a specific metal that has no adverse effect.<sup>73</sup> DI is the daily intake of leafy vegetables (kg/day), MFveg denotes the trace metal concentration in the vegetable tissues (mg/kg) and WB represents the body weight of investigated individuals. The DI is usually calculated at 0.182 kg/day for adults and 0.118 kg/day for children.<sup>14,74</sup> The body weight of investigated adults is assumed to be 55.7 kg and for children 14.2 kg.<sup>3,14</sup> If the value of HQM calculated should exceed 1 (HQM > 1), then there may be potential risk to the consumer.

## Conclusion

Trace metals are known for high mobility and bioavailability in consumed food products such as vegetables. Studies have shown that trace metals may find their way into the human system via the consumption of contaminated food crops harvested from polluted soil. The urban environment is constantly witnessing an increase in various developmental projects, with special reference to developing countries. If not properly managed, these projects may introduce contaminants, with special reference to trace metals, into the environment. However, farming activities are continuously being practised both on a small scale and a large scale around major cities and hence may be affected negatively by these contaminants.<sup>71,75</sup> Reports from the literature have suggested that leafy vegetables are capable of accumulating and storing these trace metals in their edible parts. There may be a serious problem associated with urban agriculture relating to balancing demands associated with increasing populations against potential hazards arising from the use of contaminated urban sites for food production. It is necessary to investigate and document the ability and uptake mechanism of most vegetables in order to identify and document those that can grow without accumulating trace metals in order to reduce the danger trace metals might pose to consumers. It is also important to develop new farming practices around urban city centres that will reduce or elucidate the availability and uptake of trace metals by plants.

## Acknowledgement

We thank the National Research Foundation (South Africa) for financial support.

## Authors' contributions

G.N.L. was responsible for the study under the supervision of J.O.O.

## References

1. Egziabher AG, Lee-Smith D, Maxwell DG, Memon PA, Mougeot L, Sawio C. Cities feeding people: An examination of urban agriculture in east Africa. Ottawa: International Development Research Centre; 1994.
2. Foeken D. To subsidise my income, urban farming in an east-African town. African Student Centre Series. Leiden: Brill; 2006.
3. World Health Organization (WHO). Report of the 32nd session of the Codex Committee on food additives and contaminants, ALINORM 01/12; 2000 March 20–24; Beijing, China. Geneva: WHO; 2001.
4. Vermeiren K, Adiyia B, Loopmans M, Tumwine FR, Van Rompaey A. Will urban farming survive the growth of African cities: A case study in Kampala (Uganda)? Land Use Pollut. 2013;35:40–49. <http://dx.doi.org/10.1016/j.landusepol.2013.04.012>
5. Nabulo G, Black CR, Craigon J, Young SD. Does consumption of leafy vegetables grown in peri-urban agriculture pose a risk to human health? Environ Pollut. 2012;23:389–398. <http://dx.doi.org/10.1016/j.envpol.2011.11.040>
6. Mapanda F, Mangwayana EN, Nyamangara J, Giller KE. Uptake of heavy metals by vegetables irrigated using wastewater and the subsequent risks in Harare, Zimbabwe. Phys Chem Earth ABC. 2007;32(15–18):1399–1405. <http://dx.doi.org/10.1016/j.pce.2007.07.046>
7. Ho, YB, Tai KM. Elevated levels of lead and other metals in roadside soils and grass and their use to monitor aerial metal depositions in Hong Kong. Environ Pollut J. 1988;49:37–51. [http://dx.doi.org/10.1016/0269-7491\(88\)90012-7](http://dx.doi.org/10.1016/0269-7491(88)90012-7)
8. Garcia R, Milan E. Assessment of Cd, Pb and Zn in roadside soils and grasses from Gipuzkoa (Spain). Chemosphere. 1998;37(8):1615–1625. [http://dx.doi.org/10.1016/S0045-6535\(98\)00152-0](http://dx.doi.org/10.1016/S0045-6535(98)00152-0)
9. Olowoyo JO, Van Heerden E, Fischer JL, Baker C. Trace metals in soil and leaves of *Jacaranda mimosifolia* in Tshwane area, South Africa. Atmos Environ. 2010;44:1826–1830. <http://dx.doi.org/10.1016/j.atmosenv.2010.01.048>
10. Caggiano R, Sabia S, D'Emilio M, Macchiato M, Anastasio A, Ragosta M. Metal levels in fodder, milk, dairy products, and tissues sampled in ovine farms of southern Italy. Environ Res Lett. 2005;99:48–57. <http://dx.doi.org/10.1016/j.envres.2004.11.002>
11. Gonzalez-Miqueo L, Elustondo D, Lasheras E, Santamaria JM. Use of native mosses as biomonitors of heavy metals and nitrogen deposition in the surroundings of two steel works. Chemosphere. 2010;78:965–971. <http://dx.doi.org/10.1016/j.chemosphere.2009.12.028>
12. Honour SL, Bell JNB, Ashenden TWA, Cape JN, Power SA. Responses of herbaceous plants to urban air pollution: Effects on growth, phenology and leaf surface characteristics. Environ Pollut. 2009;157:1279–1286. <http://dx.doi.org/10.1016/j.envpol.2008.11.049>
13. Tomasevic M, Vukmirovic Z, Rajsic S, Tasic M, Stevanovic B. Characterization of trace metal particles deposited on some deciduous tree leaves in an urban area. Chemosphere. 2005;61:753–760. <http://dx.doi.org/10.1016/j.chemosphere.2005.03.077>
14. Nabulo G, Black CR, Young SD. Assessing risk to human health from tropical leafy vegetables grown on contaminated urban soils. Sci Total Environ. 2010;408:5338–5351. <http://dx.doi.org/10.1016/j.scitotenv.2010.06.034>
15. Sharma RK, Agrawal M, Marshall F. Heavy metal contamination of soil and vegetables in suburban areas of Varanasi, India. Ecotox Environ Safe. 2007;66:256–288. <http://dx.doi.org/10.1016/j.ecoenv.2005.11.007>
16. Lee CW, Choi JM, Park CH. Micronutrient toxicity in seed geranium. J Am Soc Hort Sci. 1996;121(1):77–82.
17. Ali H, Khan E, Sajad MA. Phytoremediation of heavy metals – concepts and applications. Chemosphere. 2013;91:869–881. <http://dx.doi.org/10.1016/j.chemosphere.2013.01.075>



18. Li J, Li F, Liu Q. Spatial distribution and sources of dissolved trace metals in surface water of the Wei River, China. *Water Sci Technol*. 2013;67(4):817–823. <http://dx.doi.org/10.2166/wst.2012.608>
19. Lehoczky E, Szabo L, Horvath S, Marth P, Szabados I. Cadmium uptake by lettuce in different soils. *Soil Sci Plant Anal*. 1998;28:1903–1912. <http://dx.doi.org/10.1080/00103629809370080>
20. Sharma RK, Agrawal M, Marshall F. Heavy metal contamination of soil and vegetables in suburban areas of Varanasi, India. *Ecotox Environ Safe*. 2007;66:256–288. <http://dx.doi.org/10.1016/j.ecoenv.2005.11.007>
21. Nwoko CO, Mgbuehuruik L. Heavy metal contamination of ready-to-use herbal remedies in South Eastern Nigeria. *Pak J Nutr*. 2011;10:959–964. <http://dx.doi.org/10.3923/pjn.2011.959.964>
22. Birbaum K, Brogiolo R, Schellenberg M, Martinoia E, Stark WJ, Günther D. No evidence for cerium dioxide nanoparticle translocation in maize plants. *Environ Sci Technol*. 2010;44:8718–8723. <http://dx.doi.org/10.1021/es101685f>
23. Schreck E, Foucault Y, Geret F, Pradere P, Dumat C. Influence of soil ageing on bioavailability and ecotoxicity of lead carried by process waste metallic ultrafine particles. *Chemosphere*. 2011;85:1555–1562. <http://dx.doi.org/10.1016/j.chemosphere.2011.07.059>
24. Olivia SR, Espinosa EJF. Monitoring of heavy metals in topsoils, atmospheric particles and plant leaves to identify possible contamination sources. *Microchem J*. 2010;86:131–139. <http://dx.doi.org/10.1016/j.microc.2007.01.003>
25. Tomasevic M, Anicic M. Trace element content in urban tree leaves and SEM-EDAX characterisation of deposited particles. *Facta Universitatis Series: Phys Chem Technol*. 2010;8:1–13.
26. Uzu G, Sobanska S, Sarret G, Muñoz M, Dumat C. Foliar lead uptake by lettuce exposed to atmospheric fallouts. *Environ Sci Technol*. 2010;44:1036–1042. <http://dx.doi.org/10.1021/es902190u>
27. Debrowiczka T, Pirselova B, Matusikova I. The effect of cadmium on epidermis of leaves of two soybean varieties. In: Proceedings of the 13th International Scientific Conference of PhD Students, Young Scientists and Pedagogues. 2012 Sep 19–20; Nitra, Slovak Republic. Available from: <http://conferences.ukf.sk/index.php/phdconf/phdconf2012/paper/view/908>
28. Barcelo J, Poschenrieder CH. Plant water relations as affected by heavy metal stress: A review. *J Plant Nutr*. 1990;13:1–37. <http://dx.doi.org/10.1080/01904169009364057>
29. Perfus-Barbeoch L, Leonhardt N, Vavasour A, Forestier C. Heavy metal toxicity: Cadmium permeates through calcium channels and disturbs the plant water status. *Plant J*. 2002;32:539–548. <http://dx.doi.org/10.1046/j.1365-3113X.2002.01442.x>
30. Cenki Tok B, Chabaux F, Lemarchand D, Schmitt A-D, Pierret M-C, Viville D, et al. The impact of water-rock interaction and vegetation on calcium isotope fractionation in soil- and stream waters of a small forested catchment (the Strengbach case). *Geochim Cosmochim Acta*. 2009;73:2215–2228. <http://dx.doi.org/10.1016/j.gca.2009.01.023>
31. Viville MC, Bagard D, Stille ML. The impact of vegetation on REE fractionation in stream waters of a small forested catchment (the Strengbach case). *Geochim Cosmochim Acta*. 2006;70:3217–3230. <http://dx.doi.org/10.1016/j.gca.2009.01.023>
32. Zhang WF, Liu XP, Cheng HF, Zeng EY, Hu YN. Heavy metal pollution in sediments of a typical mariculture zone in South China. *Marine Poll Bull*. 2012;64:712–720. <http://dx.doi.org/10.1016/j.marpolbul.2012.01.042>
33. Sharma P, Dubey RS. Lead toxicity in plants. *Braz J Plant Physiol*. 2005;17:35–52. <http://dx.doi.org/10.1590/S1677-04202005000100004>
34. Salt DE, Prince RC, Pickering IJ, Raskin I. Mechanism of cadmium mobility and accumulation in Indian mustard. *Plant Physiol*. 1998;109:1427–1433.
35. Yang YY, Jung JY, Song WY, Suh HS, Lee Y. Identification of rice varieties with high tolerance or sensitivity to lead and characterization of the mechanism of tolerance. *Plant Physiol*. 2000;124:1019–1026. <http://dx.doi.org/10.1104/pp.124.3.1019>
36. Srivastava PK, Gupta M, Mukherjee S. Mapping spatial distribution of pollutants in groundwater of a tropical area of India using remote sensing and GIS. *Appl Geom*. 2012;4(1):21–32. <http://dx.doi.org/10.1007/s12518-011-0072-y>
37. Gabrielli R, Sanità di Toppi L. Response to cadmium in higher plants. *Environ Exp Bot*. 1999;41:105–130. [http://dx.doi.org/10.1016/S0098-8472\(98\)00058-6](http://dx.doi.org/10.1016/S0098-8472(98)00058-6)
38. Voutsas D, Grimanis A, Samara C. Trace elements in vegetables grown in industrial areas in relation to soil and air particulate matter. *Environ Pollut*. 1996;94:325–335. [http://dx.doi.org/10.1016/S0269-7491\(96\)00088-7](http://dx.doi.org/10.1016/S0269-7491(96)00088-7)
39. Lake DL, Kirk PWW, Lester JN. The fractionation characterization and speciation of heavy metals in sewage sludge and sewage amended soils. *J Environ Qual*. 1984;13:175–183. <http://dx.doi.org/10.2134/jeq1984.00472425001300020001x>
40. Scott D, Keogh JM, Allen BE. Native and low input grasses – A New Zealand high country perspective. *New Zeal J Agr Res*. 1996;39:499–512. <http://dx.doi.org/10.1080/00288233.1996.9513211>
41. Chaney RL, Malik M, Li YM, Brown SL, Angle JS, Baker AJM. Phytoremediation of soil metals. *Curr Opin Biotech*. 1997;8:279–284. [http://dx.doi.org/10.1016/S0958-1669\(97\)80004-3](http://dx.doi.org/10.1016/S0958-1669(97)80004-3)
42. Wang W, Anderson BT, Phillips N, Kaufmann RK, Potter C, Myneni RB. Feedbacks of vegetation on summertime climate variability over the North American Grasslands. Part I: Statistical analysis. *Earth Interact*. 2006;10:1–27. <http://dx.doi.org/10.1175/EI196.1>
43. Tangahu BV, Abdullah SRS, Basri H, Idris M, Anuar N, Mukhlis M. A review on heavy metals (As, Pb and Hg) uptake by plants through phytoremediation. *Int J Chem Eng*. 2011;2011, Art. #939161, 31 pages. <http://dx.doi.org/10.1155/2011/939161>.
44. Gregson SK, Alloway BJ. Gel permeation chromatography studies on the speciation of lead in solutions of heavily polluted soils. *J Soil Sci*. 1984;35:55–61.
45. Fijalkowski K, Kacprzac M, Grobelak A, Placek A. The influence of selected soil parameters on the mobility of heavy metals in soils. *Inz I Ochr Srodo*. 2012;15:81–92.
46. Clemente EP, Schaefer CEGR, Novais RF, Viana JH, Barros NF. Soil compaction around *Eucalyptus grandis* roots: A micromorphological study. *Soil Res*. 2005;43:139–146. <http://dx.doi.org/10.1071/SR04069>
47. Clemente R, Walker DJ, Roig A, Bernal MP. Heavy metal bioavailability in a soil affected by mineral sulphides contamination following the mine spillage at Aznalcollar (Spain). *Biodegradation*. 2003;14:199–205. <http://dx.doi.org/10.1023/A:1024288505979>
48. Walker TR, Young SD, Crittenden PD, Zhang H. Anthropogenic metal enrichment of snow and soil in north-eastern European Russia. *Environ Pollut*. 2003;121:11–21. [http://dx.doi.org/10.1016/S0269-7491\(02\)00212-9](http://dx.doi.org/10.1016/S0269-7491(02)00212-9)
49. Schulman LJ, Sargent EV, Naumann BD, Faria EC, Dolan DG, Wargo JP. A human health risk assessment of pharmaceuticals in the aquatic environment. *Hum Ecol Risk Assess*. 2002;8(4):657–680. <http://dx.doi.org/10.1080/10807030290879899>
50. Lasat MM. The use of plants for the removal of toxic metals from contaminated soil. Washington, DC: American Association for the Advancement of Environmental Science and Engineering; 2012. p. 1–33.
51. Jung MC. Heavy metal contamination in soils and factors affecting metal uptake by plants in the vicinity of a Korean Cu-W mine. *Sensors*. 2008;8:2413–2423. <http://dx.doi.org/10.3390/s8042413>
52. Corey RB, King LD, Lue-Hing C, Fanning DS, Street JJ, Walker JM. Effects of sludge properties on accumulation of trace elements by crops. In: Page AL, Logan TJ, Ryan JA, editors. Land application of sludge, food chain implications. Chelsea: Lewis Publications; 1987.
53. Barber SA. Soil nutrient bioavailability: A mechanistic approach. New York: John Wiley; 1984.
54. Tomašević M, Aničić M. Trace element content in urban tree leaves and SEM-EDAX characterization of deposited particles. *Facta Universitatis Series: Phys Chem Technol*. 2010;8(1):1–13.
55. Yadav SK. Heavy metals toxicity in plants: An overview on the role of glutathione and phytochelatins in heavy metal stress tolerance of plants. *S Afr J Bot*. 2010;76:167–179. <http://dx.doi.org/10.1016/j.sajb.2009.10.007>
56. Lasat MM, Fuhrmann M, Ebbs SD, Cornish JE, Kochvian LV. Phytoremediation of a radiocesium-contaminated soil: Evaluation of cesium-137 bioaccumulation in the shoots of three plant species. *J Environ Qual*. 1998;27:165–169. <http://dx.doi.org/10.2134/jeq1998.271165x>

57. Verkleij JAC, Golan-Goldhirsh A, Antosiewicz DM, Schwitzguebel J, Schroder P. Dualities in plant tolerance to pollutants and their uptake and translocation to the upper plant parts. *Environ Exp Bot.* 2009;67(1):10–22. <http://dx.doi.org/10.1016/j.envexpbot.2009.05.009>
58. Baker AJM. Meta tolerance. *New Phytol.* 1987;106:93–111. <http://dx.doi.org/10.1111/j.1469-8137.1987.tb04685.x>
59. Hamadouche NA, Aoumeur H, Djedjai S, Slimani M, Aoues A. Phytoremediation potential of *Raphanus sativus* L. for lead contaminated soil. *Act Biol Szeg.* 2012;56(1):43–49.
60. Krumins JA, Long ZT, Steiner CF, Morin PJ. Indirect effects of food web diversity and productivity on bacterial community function and composition. *Funct Ecol.* 2006;20:514–521. <http://dx.doi.org/10.1111/j.1365-2435.2006.01117.x>
61. Frey B, Stemmer M, Widmer F, Luster J, Sperisen C. Microbial activity and community structure of a soil after heavy metal contamination in a model forest ecosystem. *Soil Biol Biochem.* 2006;38:1745–1756. <http://dx.doi.org/10.1016/j.soilbio.2005.11.032>
62. Dahmani-Muller H, Oort FV, Denaix L. Is metal extraction by *Arabidopsis halleri* related to exchangeable metal rates in soils amended with different metal-bearing solids? *Environ Pollut.* 2002;117(3):487–498. [http://dx.doi.org/10.1016/S0269-7491\(01\)00180-4](http://dx.doi.org/10.1016/S0269-7491(01)00180-4)
63. Lone MI, He ZL, Stoffella PJ, Yang XE. Phytoremediation of heavy metal polluted soils and water: Progresses and perspectives. *J Zhejiang Univ-Sc B.* 2008;9:210–220. <http://dx.doi.org/10.1631/jzus.B0710633>
64. Mellem JJ, Baijnath H, Odhav B. Bioaccumulation of Cr, Hg, As, Pb, Cu and Ni with the ability for hyperaccumulation by *Amaranthus dubius*. *Afr J Agric Res.* 2008;7:591–596.
65. Olowoyo JO, Okedeyi NM, Mkolo NM, Lion GN, Mdakane STR. Uptake and translocation of heavy metals by medicinal plants growing around a waste dump site in Pretoria, South Africa. *S Afr J Bot.* 2011;78:116–121. <http://dx.doi.org/10.1016/j.sajb.2011.05.010>
66. Poschenrieder C, Tolra R, Barcelo J. Can metals defend plants against biotic stress? *Trends Plant Sci.* 2006;11:288–295. <http://dx.doi.org/10.1016/j.tplants.2006.04.007>
67. Miteva E, Hristova D, Nenova V, Manava S. Arsenic as a factor affecting virus infection in tomato plants; changes in plant growth, peroxidase activity and chloroplast pigments. *Sci Hortic-Amsterdam.* 2005;105:343–358. <http://dx.doi.org/10.1016/j.scienta.2005.01.026>
68. Yadav SK. Heavy metal toxicity in plants. An overview of the role of glutathione and phytochelatins in heavy metals stress tolerance of plants. *S Afr J Bot.* 2010;76:167–179. <http://dx.doi.org/10.1016/j.sajb.2009.10.007>
69. Hu H. Human health and heavy metal exposure. In: McCally M, editor. *Life support. The environment and human health.* Cambridge, MA: Massachusetts Institute of Technology Press; 2012. p. 1–12.
70. Shen L, Xia B, Dai X. Residues of persistent organic pollutants in frequently-consumed vegetables and assessment of human health risk based on consumption of vegetables in Huizhou, South China. *Chemosphere.* 2013;93:2254–2263. <http://dx.doi.org/10.1016/j.chemosphere.2013.07.079>
71. Xu D, Zhou P, Zhan J, Gao Y, Dou C, Sun Q. Assessment of trace metal bioavailability in garden soils and health risks via consumption of vegetables in the vicinity of Tongling mining area, China. *Ecotox Environ Safe.* 2013;90:103–111. <http://dx.doi.org/10.1016/j.ecoenv.2012.12.018>
72. Olowoyo JO, Lion GN. Uptake and translocation of trace metals from soil collected around waste dump sites and a mining area by vegetables [MSc dissertation]. Pretoria: University of Limpopo; 2013.
73. Martin S, Griswold W. Human health effects of heavy metals. *Environ Sci Technol Briefs Citizens.* 2009;15:1–6.
74. Mahmood A, Malik RN. Human health risk assessment of heavy metals via consumption of contaminated vegetables collected from different irrigation sources in Lahore, Pakistan. *Arab J Chem.* 2014;7:91–99. <http://dx.doi.org/10.1016/j.arabjc.2013.07.002>
75. Lei M, Yue QL, Chen TB, Huang ZC, Liao XY, Liu YR, et al. Heavy metal concentrations in soils and plants around Shizhuyuan mining area of Hunan Province. *Acta Ecol Sin.* 2005;25(5):1146–1151. In Chinese with English abstract.



# Neutrophil extracellular traps and their role in health and disease

## AUTHORS:

Jan G. Nel<sup>1</sup>  
Annette J. Theron<sup>2</sup>  
Roger Pool<sup>1</sup>  
Chrisna Durandt<sup>3</sup>  
Gregory R. Tintinger<sup>4</sup>  
Ronald Anderson<sup>3</sup>

## AFFILIATIONS:

<sup>1</sup>Department of Haematology, University of Pretoria, Pretoria, South Africa

<sup>2</sup>Department of Immunology, University of Pretoria, Pretoria, South Africa

<sup>3</sup>Institute for Cellular and Molecular Medicine, SAMRC Extramural Unit for Stem Cell Research and Therapy, Department of Immunology, University of Pretoria, Pretoria, South Africa

<sup>4</sup>Department of Internal Medicine, University of Pretoria and Steve Biko Academic Hospital, Pretoria, South Africa

## CORRESPONDENCE TO:

Jan Nel

## EMAIL:

Jan.Nel@up.ac.za

## POSTAL ADDRESS:

Department of Haematology, Faculty of Health Sciences, University of Pretoria, Private Bag X323, Arcadia 0007, South Africa

## DATES:

Received: 20 Feb. 2015

Revised: 08 June 2015

Accepted: 18 June 2015

## KEYWORDS:

antimicrobial activity; autoimmune disease; innate immunity; phagocytes; reactive oxygen species

## HOW TO CITE:

Nel JG, Theron AJ, Pool R, Durandt C, Tintinger GR, Anderson R. Neutrophil extracellular traps and their role in health and disease. *S Afr J Sci.* 2016;112(1/2), Art. #2015-0072, 9 pages. <http://dx.doi.org/10.17159/sajs.2016/20150072>

© 2016. The Author(s).  
Published under a Creative Commons Attribution Licence.

The human innate immune system is indispensable for protection against potentially invasive microbial and viral pathogens, either neutralising them or containing their spread until effective mobilisation of the slower, adaptive (specific), immune response. Until fairly recently, it was believed that the human innate immune system possessed minimal discriminatory activity in the setting of a rather limited range of microbicidal or virucidal mechanisms. However, recent discoveries have revealed that the innate immune system possesses an array of novel pathogen recognition mechanisms, as well as a resourceful and effective alternative mechanism of phagocyte (predominantly neutrophil)-mediated, anti-infective activity known as NETosis. The process of NETosis involves an unusual type of programmed, purposeful cell death, resulting in the extracellular release of a web of chromatin heavily impregnated with antimicrobial proteins. These structures, known as neutrophil extracellular traps (NETs), immobilise and contribute to the eradication of microbial pathogens, ensuring that the anti-infective potential of neutrophils is sustained beyond the lifespan of these cells. The current review is focused on the mechanisms of NETosis and the role of this process in host defence. Other topics reviewed include the potential threats to human health posed by poorly controlled, excessive formation of NETs, specifically in relation to development of autoimmune and cardiovascular diseases, as well as exacerbation of acute and chronic inflammatory disorders of the airways.

## Introduction

Until fairly recently, the protective activities of the human innate immune system, which are highly conserved throughout evolution, were thought to be achieved by a limited range of mechanisms with minimal discriminatory potential. Predominant amongst these mechanisms are the physical barriers presented by the skin and mucous membranes, engulfment and destruction of microbial and viral pathogens by resident and migratory phagocytes, and the non-specific antimicrobial activity of various blood and tissue proteins. These mechanisms either prevent infection or, in the case of a breach by a pathogen, contain the infection until adaptive (specific) host defences are effectively mobilised. Over the last decade, however, a number of significant discoveries have revealed that the human innate immune system not only possesses a level of discrimination previously considered improbable, but also includes additional, resourceful mechanisms of phagocyte-mediated antimicrobial and antiviral activity.

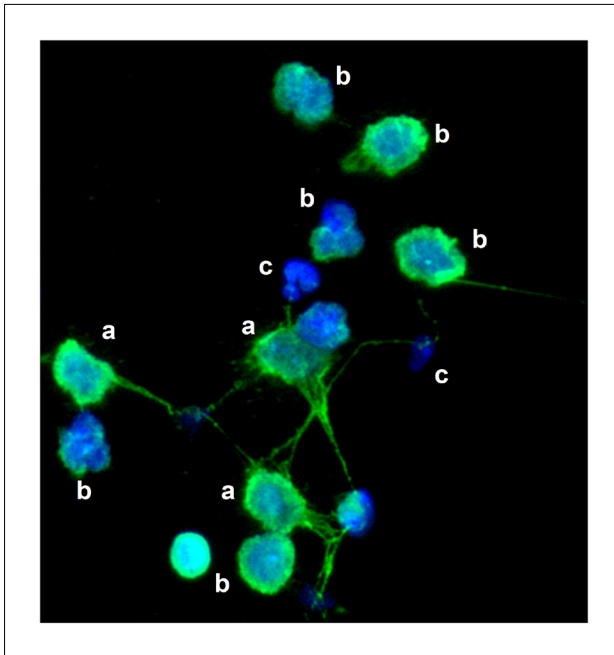
With respect to pathogen detection, cells of the innate immune system (phagocytes, mast cells, basophils and dendritic cells), as well as epithelial cells, have been found to possess various types of pathogen recognition receptors which recognise conserved molecular structures broadly expressed on or in microbial and viral pathogens. These receptors include the Toll-like receptors, the nucleotide oligomerisation domain-like receptors, and the abundant cytosolic microbial and viral nucleic acid sensors, activation of which initiates a potentially protective inflammatory response. These receptors have been the subject of several recent reviews.<sup>1,2</sup>

In 2004, Brinkmann et al.<sup>3</sup> described an unusual mechanism by which human blood neutrophils immobilise pathogens extracellularly, exposing them to a highly concentrated array of anti-infective proteins. Neutrophils (also known as granulocytes or polymorphonuclear leucocytes) are the predominant small circulating phagocytes. These cells have an estimated lifespan of 5.4 days in the circulation,<sup>4</sup> which is longer when they are exposed to anti-apoptotic cytokines. These cells exit the circulation via transendothelial migration and chemotaxis to sites of microbial and viral infection where they phagocytose and destroy pathogens via intracellular exposure to microbicidal and virucidal reactive oxygen species (ROS), proteases and proteins.

The studies of Brinkmann et al.<sup>3</sup> enhanced the body of knowledge on neutrophil function through the discovery that these cells also respond to infectious challenges via the formation of neutrophil extracellular traps (NETs). NETs are web-like structures composed of decondensed chromatin heavily impregnated with different antimicrobial granular proteins which capture, neutralise and kill a variety of pathogens. NETs are produced predominantly by neutrophils, but also by other cell types of the innate immune system such as monocytes and macrophages, eosinophils, basophils and mast cells, in which the process is termed ETosis. Phylogenetic studies have revealed that ETosis is a highly conserved 'ancient defence weapon, predating the evolution of the coelom' operative in haemocytes, the phagocytic cells of invertebrates.<sup>5</sup>

NETs form large extracellular barriers to bacterial dissemination, and provide a mechanism for localised concentration of effector molecules. Importantly, NET formation has been demonstrated in both the clinical and experimental infection settings using immunohistochemistry and spinning disc vital microscopy.<sup>6</sup> Several strategies to measure NET formation *in vitro* have also been described, including immunofluorescence and electron microscopic procedures, as well as spectrofluorimetric and other methods which detect extracellular DNA and associated granule proteins.<sup>3</sup> Human neutrophils undergoing NETosis *in vitro* are shown in Figure 1.

NETs are the topic of this review, which is focused on mechanisms of NETosis and the role of this process in host defence, as well as on the potentially harmful consequences of excessive NETosis for the host and possible pharmacological control strategies.



**Figure 1:** Fluorescence micrograph showing isolated human neutrophils undergoing NETosis. Following activation with phorbol myristate acetate (6.25 ng/mL), a potent stimulator of nicotinamide adenine dinucleotide phosphate (NADPH) oxidase, and the production of reactive oxygen species, adherent neutrophils were exposed, sequentially, to unlabelled polyclonal rabbit antibodies, to citrullinated histone H3 residues and to Alexa Fluor 488 labelled goat anti-rabbit antibodies, before DNA was stained with a nuclear dye (DAPI). DNA is stained blue, while neutrophil extracellular traps (NETs), containing both DNA and citrullinated histone residues, are stained green. Neutrophils which have undergone NETosis, cells in the early stages of NETosis, and a non-NETotic neutrophil showing intact, multi-lobed nuclear morphology (stained blue), are labelled a, b and c, respectively.

## Mechanisms of NET formation

Although the exact molecular and biochemical mechanisms involved in the formation and release of NETs are incompletely understood, three different types of NETotic pathway have been described. The best characterised of these pathways is slow and leads to lytic cell death over the course of 2–3 h. The second pathway has been described as a rapid mechanism (vital), independent of cell lysis, which requires the rapid (within minutes) vesicular release of neutrophil nuclear contents.<sup>6,7</sup> Unlike the first two mechanisms of NETosis which involve release of nuclear DNA, the third mechanism, which is also non-lytic, involves the release of mitochondrial DNA.<sup>8</sup>

Lytic NETosis, also known as suicidal NETosis, is an active cell death related process distinct from either necrosis or apoptosis.<sup>9</sup> This process requires chromatin decondensation, nuclear envelope disintegration and a mixing of nucleic acids and granule proteins within an intracellular vacuole. Subsequent release of vacuole contents into the extracellular domain is preceded by plasma membrane perforation or lysis. Various stages in the process constituting lytic NETosis have been identified. The interaction of pathogens and their products with their counter-receptors on or in neutrophils includes, in addition to the pathogen recognition receptors mentioned above, the opsonin receptors FcR (receptor for pathogen-bound immunoglobulin G) and CR3 (receptors for pathogen-bound complement components C3b and C3bi), which promote adherence of pathogens to neutrophils. This interaction, in turn, leads to a series of pathogen-activated intracellular signalling events, most importantly: (1) activation of the receptor-linked signalling complex protein kinase C(PKC)-raf/-MEK-ERK which mediates activation of the neutrophil membrane-associated, electron-transporting, ROS-generating

system, NADPH oxidase<sup>10</sup>; (2) ROS-mediated activation of intracellular signalling pathways converging on the cytosolic transcription factor, nuclear factor kappa B (NFκB), which, following nuclear translocation, has been proposed to promote NETosis by initiating transcription of the peptidylarginine deiminase 4 (PAD4) gene, as well as by blocking apoptosis, an alternative pathway of programmed cell death<sup>11</sup>; and (3) receptor-mediated increases in cytosolic Ca<sup>2+</sup> via activation of phospholipase C, which, in turn, leads to Ca<sup>2+</sup>-dependent activation of PAD4. ROS also promote collapse of both the cytoplasmic granule and nuclear membranes, enabling access of cytoplasmic PAD4 and granule proteins to chromatin. PAD4 mediates histone hypercitrullination (conversion of protein arginine residues to citrulline), a key event in chromatin decondensation<sup>12,13</sup>, which is facilitated by the limited proteolysis of nuclear histones mediated by granule enzymes – neutrophil elastase and myeloperoxidase (MPO) – operating in unison<sup>3,14</sup>. While ROS appear to promote the rupture of both the cytosolic granule and nuclear membranes, a recent study has also implicated the involvement of the granule antimicrobial polypeptide LL-37 (which consists of 2 N-terminal leucines and a total of 37 amino acids).<sup>15</sup> LL-37 is a cationic amphiphilic polypeptide of the cathelicidin family which binds to anionic membrane phospholipids promoting membrane disruption. The nuclear membrane appears particularly vulnerable to the disruptive actions of LL-37. In addition, the cationic properties of LL-37 also promote binding of the polypeptide to neutrophil DNA, increasing resistance to degradation of NETs by microbial nucleases.<sup>16</sup>

The proposed mechanism of lytic NET formation based on current knowledge is shown in Figure 2.

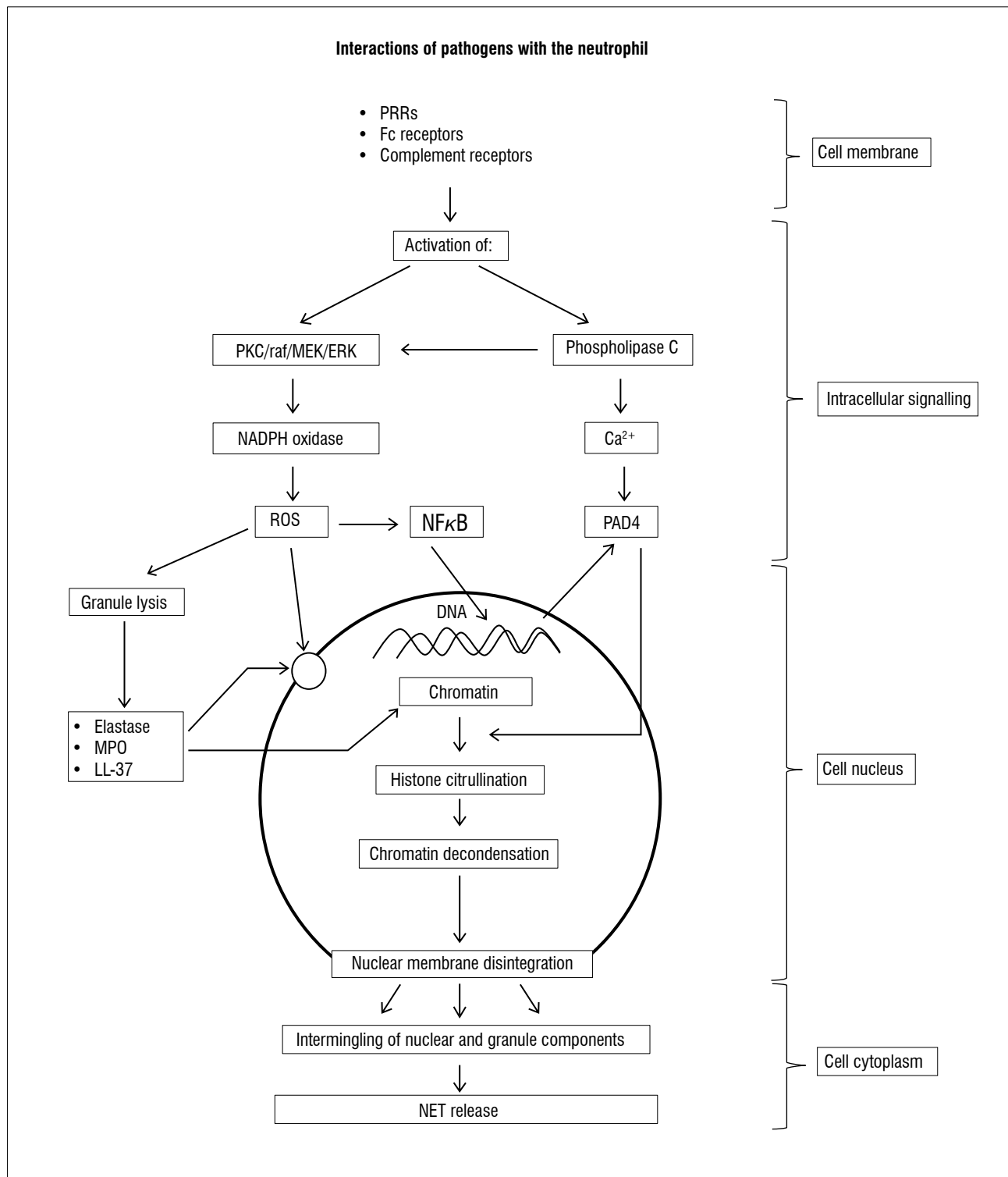
The concept that alternative NETosis pathways exist, in addition to the lytic-cell death pathway, has been advanced by several groups.<sup>6,8,17,18</sup> One such pathway is vital NETosis. Vital NETosis is a rapid process whereby cell viability and function are retained in the context of controlled, incremental discharge of nuclear material, following exposure of neutrophils to various microorganisms and their products, and appears to be a generalised response against various classes of microbial pathogens.<sup>18,19</sup> In this setting, neutrophils are stimulated to ‘release NETs via nuclear envelope bleb formation and vesicular exportation, preserving the integrity of the plasma membrane’.<sup>6,20</sup> Importantly, and also in contradistinction to lytic NETosis, the requirement for involvement of NADPH oxidase in vital NETosis is variable, with increases in cytosolic calcium seemingly adequate in the case of some stimuli such as bacterial pore-forming toxins<sup>6</sup> and calcium ionophores<sup>7</sup>, or alternatively via ROS-independent activation of NFκB<sup>21</sup>. Unlike lytic NETosis, only 20–25% of the neutrophil population undergoes NETosis on exposure to microbial pathogens or their products *in vitro*.<sup>20</sup> This observation is not only consistent with the existence of a sub-population of neutrophils highly specialised for the performance of vital NETosis, but also suggests that vital, as opposed to lytic, NETosis is the more physiologically relevant of the two processes.<sup>20</sup>

The formation of mitochondrial-derived NETs is also a type of vital NETosis with variable dependence on activation of NADPH oxidase according to the nature of the cell activator. Although the existence of mitochondrial DNA-containing NETs has been demonstrated *in vitro*, less is known about its role in host defence which is likely to be limited by the absence of histones.<sup>22,23</sup> Interestingly, mitochondrial DNA has been found to induce NET formation, consistent with a role in the amplification of NETosis.<sup>22</sup>

## NET constituents

The DNA scaffolding of NETs is provided by decondensed nuclear chromatin. A variety of NET-associated proteins, predominantly highly negatively charged histones, is arranged on this scaffolding. Via their strong positive charge, histones facilitate the adhesion to, as well as the sequestration of, microbial and viral pathogens<sup>24</sup> and also confer bactericidal activity, as does DNA.<sup>25</sup> The remaining NET-associated proteins comprise granule, cytoplasmic and cytoskeletal proteins, as well as metabolic enzymes.<sup>24</sup> The concept of a core NET-associated proteome that remains constant regardless of the specific agent responsible for NET induction, has been proposed by Rahman and Gadjeva<sup>26</sup>. The core





**Figure 2:** Basic mechanism of lytic NETosis. Exposure of neutrophils to pathogens or their cell-wall and intracellular components bound to pathogen recognition receptors (PRRs) or receptors for antibody (Fc) or complement components is linked to activation of: (1) the protein kinase C (PKC)/raf/erk kinase (MEK)/extracellular signal regulated kinase (ERK) intracellular signalling axis, which, in turn, activates the membrane-bound, electron-transporting, reactive oxygen species (ROS)-generating complex, nicotinamide adenine dinucleotide phosphate (NADPH) oxidase; and (2) the enzyme phospholipase C which cleaves membrane phosphatidylinositol to generate diacylglycerol (which also activates NADPH oxidase) and inositol triphosphate, which mobilises calcium ( $Ca^{2+}$ ) from intracellular stores. ROS initiate activation of the latent cytosolic transcription factor, nuclear factor kappa B (NFkB), which translocates to the nucleus and induces, amongst others, the gene encoding the  $Ca^{2+}$ -dependent, pro-NETotic enzyme peptidylarginine deiminase 4 (PAD4). ROS also promote disruption of the membranes of cytosolic granules leading to release of granule antimicrobial proteins and enzymes, including LL-37, which act in concert with ROS to augment nuclear membrane disintegration, while elastase and myeloperoxidase (MPO) also mediate chromatin decondensation. Nuclear membrane disruption enables PAD4 to access the nucleus, which, in turn, triggers the series of events culminating in neutrophil extracellular trap (NET) formation. This involves PAD4-mediated citrullination of nuclear histones followed by chromatin decondensation, and intermingling of nuclear and granule antimicrobial components in the cytoplasm to form a matrix which is released extracellularly as NETs.

**Table 1:** A summary of the major antimicrobial constituents of neutrophil extracellular traps

Constituent	Origin	Anti-infective spectrum
DNA	Cell nucleus	Broad-spectrum activity
Histones	Cell nucleus	Broad-spectrum activity
Myeloperoxidase (MPO)	Neutrophil primary granules	Broad-spectrum when combined with hydrogen peroxide
Neutrophil elastase (NE)	Neutrophil primary granules	Serine protease with broad-spectrum activity
Proteinase 3 (PR3)	Neutrophil primary granules	Serine proteinase with broad-spectrum activity
Cathepsin G	Neutrophil primary granules	Serine proteinase with broad-spectrum activity
$\alpha$ -Defensins	Neutrophil primary granules	Broad-spectrum activity; also known as human neutrophil peptides 1–4 (Hnp 1–4)
Azurocidin (cationic antimicrobial peptide-37)	Neutrophil primary granules	Broad-spectrum antimicrobial polypeptide, also known as cationic antimicrobial peptide-37
Bactericidal permeability-increasing protein (BPI)	Neutrophil primary granules	Selective activity against Gram-negative bacterial pathogens
Lysozyme	Neutrophil primary, secondary and tertiary granules	Selective activity against Gram-positive bacterial pathogens
LL-37 (cathelicidin)	Neutrophil secondary granules	Broad-spectrum activity
Lactoferrin	Neutrophil secondary granules	Broad-spectrum activity
Calprotectin	Neutrophil cytosol	Selective activity against the yeast <i>Candida albicans</i> and <i>Aspergillus</i> fungal species

Sources<sup>16,23-25,37</sup>

NET-associated proteome is made up of a combination of 19 constituent proteins, arranged in association with a variety of decorative proteins. The fact that the protein structure of NETs exhibits a constant element to its composition supports the hypothesis that NET formation is indeed an innate immune system response which is non-specific.<sup>26</sup> The major antimicrobial components of NETs are summarised in Table 1.

## Beneficial effects of NETosis

### Role in host defence

NETs have been shown to degrade microbial and viral virulence factors and to restrict pathogens by forming a physical barrier that prevents dissemination. Examples of bacterial, fungal and protozoal parasitic pathogens which are ensnared in NETs are shown in Table 2. While some bacterial pathogens such as *Pseudomonas aeruginosa*<sup>27</sup>, *Borrelia burgdorferi*<sup>28</sup> and *Burkholderia pseudomallei*<sup>29</sup> are killed following entrapment in NETs, others appear less vulnerable, probably as a result of the production of anti-adhesive surface structures such as polysaccharide capsules and/or NET-degrading nucleases, as indicated in Table 2.<sup>24,27-47</sup>

Neutrophils appear to be particularly adept at sensing microbial size, with large microorganisms the most effective inducers of NETosis.<sup>31</sup> For example, in the case of *Candida albicans*, the yeast responsible for most fungal infections in humans, NETs are effectively induced by this pathogen in both the yeast and hyphal forms.<sup>48</sup> Protozoa such as *Toxoplasma gondii*, *Plasmodium falciparum* and *Leishmania* spp. have also been shown to possess the requisite signals to trigger NET formation, which, in some cases, may lead to the death of entrapped

parasites, as well as interference with the invasion of host cells. Some microbial pathogens, such as the opportunistic Gram-negative bacterium *Acinetobacter baumannii* and the fungus *Cryptococcus neoformans*, do not appear to activate NETosis, which in the case of the latter has been attributed predominantly to the ability of the polysaccharide capsule to prevent the requisite signalling mechanisms.<sup>49,50</sup>

Based on observations of pathogens such as *Staphylococcus aureus* and *Streptococcus pneumoniae*, which withstand NETs, it has been proposed that the primary function of NETs is to immobilise, weaken and expose infective agents to other antimicrobial cellular and humoral components of the innate immune system, including tissue macrophages and the proteins of the complement system respectively.<sup>51-53</sup> An example of this type of cooperation involves the interaction of neutrophils and macrophages in the eradication of the bacterial pathogen *Mycobacterium tuberculosis*. This pathogen has also been reported to induce NETs in which it is trapped extracellularly, but remains viable. Entrapped bacilli are then engulfed by alveolar macrophages, the primary cell type involved in the eradication of *M. tuberculosis*.<sup>30</sup>

With respect to the role of NETs in antiviral host defence, neutrophils have been shown to detect HIV-1 via interaction with pathogen recognition receptors which recognise viral RNA. This detection in turn triggers NET formation, leading to NET-mediated inactivation of HIV-1, resulting from exposure to MPO-derived oxidants and  $\alpha$ -defensins.<sup>54</sup> This response may, however, be attenuated via the release of the anti-inflammatory cytokine, interleukin-10, from bystander HIV-infected dendritic cells, which, in turn, inhibits NET formation.<sup>54</sup>

**Table 2:** Examples of microbial pathogens which induce NETosis and their escape strategies

Pathogen	Type	Protective actions of neutrophil extracellular traps (NETs)	Escape/survival mechanisms
<i>Pseudomonas aeruginosa</i>	Bacterium	Entrapment and killing of some strains, others escape	Shedding of outer membrane vesicles which compete with NET binding sites, as well as acquired resistance
<i>Borrelia burgdorferi</i>	Bacterium	Entrapment and killing	None evident
<i>Burkholderia pseudomallei</i>	Bacterium	Entrapment and killing of strains with low capsule expression	Evasion by capsular polysaccharides
<i>Staphylococcus aureus</i>	Bacterium	Entrapment only	Subversion of macrophage-mediated uptake and killing of NET-associated bacteria by production of pro-apoptotic deoxyadenosine
<i>Streptococcus pneumoniae</i> and other streptococci species	Bacterium	Entrapment of strains with low capsule expression	Capsule polysaccharide-mediated interference with binding to NETs and escape as a result of production of microbial nucleases
<i>Neisseria meningitidis</i>	Bacterium	Entrapment resulting in decreased proliferation	Shedding of competitive outer membrane vesicles and adaptive cell surface modifications
<i>Neisseria gonorrhoea</i>	Bacterium	Entrapment and limited bactericidal activity	Escape as a result of subsequent production of a thermonuclease and inhibition of production of pro-NETotic reactive oxygen species by neutrophils
<i>Vibrio cholera</i>	Bacterium	Entrapment and limited bactericidal activity	Escape as a result of subsequent production of two microbial nucleases
<i>Mycobacterium tuberculosis</i>	Bacterium	Entrapment without killing	Resistance to killing mediated by the thick, waxy outer coat of the pathogen
<i>Candida albicans</i>	Yeast	Entrapment and killing	Escape reported, but mechanisms unknown
<i>Aspergillus</i> species	Fungus	Entrapment and killing	No escape mechanisms described
<i>Leishmania</i> species	Protozoal parasite	Entrapment and limited killing	Escape mediated by release of 3'-nucleotidase/nuclease activity
<i>Toxoplasma gondii</i>	Protozoal parasite	Entrapment and killing	Escape mechanisms not yet identified
<i>Plasmodium falciparum</i>	Protozoal parasite	Entrapment and limited killing	Escape mechanisms not yet identified

Sources<sup>24,27-47</sup>

Although the exact role of NETs in the host response to infection remains to be convincingly elucidated, the increased susceptibility of patients with chronic granulomatous disease (CGD) to infection is noteworthy in this context. This condition is an inherited primary immunodeficiency disorder caused by a complete absence of NADPH oxidase and failure of phagocytes to generate ROS, and consequently NETs. The severe impairment of neutrophil protective activity as a result of the combined absence of production of antimicrobial ROS and NETs causes the patients to suffer from severe and often life-threatening infections.<sup>9</sup> However, NET production by neutrophils in chronic granulomatous disease is responsive to activation by ROS-independent pro-NETotic mechanisms.<sup>21</sup> In addition, transient and acquired abnormalities of NET formation have also been demonstrated in human neonates and the elderly, respectively – a previously unrecognised deficit in extracellular bacterial killing which may underpin age-associated vulnerability to microbial and viral infection.<sup>55,56</sup>

In addition, the increased susceptibility of humans with stable or transient severe neutropenia for development of disseminated fungal infections has been attributed to attenuation of 'trapping' by NETs.<sup>20</sup>

### *Evasion of NETs by pathogens*

Several strategies have been described that enable pathogens to evade NET-mediated immobilisation and/or killing by NET-associated proteins, many of which are listed in Table 2. These strategies include the production of nuclease enzymes that degrade the DNA backbone of the NET structures by various types of microbial pathogens, including, but not limited to, pathogens of the *Streptococcus* and *Staphylococcus* genera. In addition, the acquisition by microorganisms and viruses of molecular patterns that interfere with pathogen recognition by pathogen recognition receptors has also been reported to attenuate NET formation.<sup>34-36</sup> The acquisition of a cell capsule, for example, is of particular benefit in evading NET-mediated trapping of organisms, by altering the surface charge of bacteria to neutral and thereby negating the electrostatic attraction posed by positively charged NET fibres and histone residues.<sup>36</sup>

### *The involvement of NETs in the pathophysiology of disease*

Despite beneficial effects in host defence, NETosis may occur at the expense of injury to the host.<sup>57</sup> Inappropriate and/or excessive NET

formation has been documented in the following autoimmune, cardiovascular and pulmonary diseases.

### *The role of NETs in autoimmune diseases*

#### Systemic lupus erythematosus

Systemic lupus erythematosus (SLE) is a complex and heterogeneous disease, with patients displaying a variety of symptoms of which glomerulonephritis is particularly serious.<sup>57,58</sup> This condition has a definite female preponderance and a prevalence which seemingly varies according to race. The hallmark of SLE is the overproduction of autoantibodies against a range of nuclear antigens including, not only DNA and histones, but also neutrophil granule proteins.<sup>57</sup> These autoantibodies are believed to contribute significantly to disease pathogenesis. In this context, several studies have reported that the ability to degrade NETs was reduced in a subset of patients with a severe form of SLE which was associated with both glomerulonephritis and the presence of circulating autoantibodies reactive with various constituents of NETs.<sup>59,60</sup> Mechanistically, disassembly of NETs in the physiological setting is mediated by the enzyme serum endonuclease DNase1, interference with which is likely to favour persistence and exaggerated immunogenicity of NETs.<sup>60</sup> In this setting, binding of anti-NET antibodies to NETs has been reported to prevent access of DNase I to NETs. The consequence is impairment of DNase1 function, resulting in failure to dismantle NETs, correlating with renal involvement in SLE.<sup>60</sup> As alluded to earlier, it appears that NETosis in SLE also involves a subset of neutrophils known as low density granulocytes.<sup>61</sup>

#### Rheumatoid arthritis

Rheumatoid arthritis (RA) is a chronic, systemic autoinflammatory disease which primarily affects the synovial joints. It occurs at high frequency (1–3%) in the general population with a female preponderance and, like SLE, is associated with high morbidity and mortality. The majority of RA patients present with high levels of circulating antibodies to citrullinated proteins known as anti-citrullinated peptide antibodies (ACPAs) which are serodiagnostic for RA. Although a clear mechanistic relationship between dysregulation of NETosis and production of ACPAs remains to be established, it is noteworthy that neutrophils from patients with RA exhibit exaggerated NETosis in the circulation, the skin and rheumatoid joint when compared with neutrophils from healthy controls and patients with osteoarthritis.<sup>62,63</sup> Other supporting evidence includes: (1) the finding of a significant positive correlation between NET formation and serum levels of anti-citrullinated peptide antibodies, as well as with other circulating biomarkers of inflammation and neutrophil activation; (2) a report that citrullinated histone 4, a component of NETs, is reactive with ACPAs; and (3) the finding that NETs act as strong stimulants of fibroblast-like synoviocytes (cells that invade cartilage in RA).<sup>62</sup> Taken together, these findings appear to implicate dysregulation of NETosis in the pathogenesis of RA which, in turn, may lead to the identification of novel targets for the treatment of this and other diseases.<sup>62</sup>

#### Small vessel vasculitis

Small vessel vasculitis is a chronic autoinflammatory condition in which small blood vessels show necrotic inflammation. The condition is associated with the presence of anti-neutrophil cytoplasmic autoantibodies (ANCA). The main targets for ANCA are the granule enzymes MPO and PR3 (proteinase 3).<sup>64</sup> Kessenbrock et al.<sup>65</sup> observed that the binding of ANCA to neutrophils resulted in activation of NETosis. They also demonstrated typical components of NETs present in kidney biopsies of patients with small vessel vasculitis. The NETs were decorated with the autoantigens MPO and PR3. Deposition of NETs in inflamed kidneys suggests that NET formation plays a pathogenic role in autoimmune small vessel vasculitis by presenting autoantigens to the immune system with resultant vascular damage.<sup>65,66</sup>

### *The role of NETs in deep vein thrombosis*

Deep vein thrombosis (DVT) is the formation of a blood clot (or thrombus) in a deep vein, predominantly in the legs. DVT can be triggered by disturbances in venous blood flow, activation or dysfunction

of the vascular endothelium, and hypercoagulability. NETs provide a new link between innate immunity and hypercoagulability,<sup>67</sup> stimulating the coagulation process by activating platelets, the coagulation cascade and the vascular endothelium.<sup>68</sup> NETs provide a scaffold for platelet and red blood cell adhesion and also concentrate effector proteins involved in thrombosis.<sup>68</sup> It is speculated that NET-associated enzymes may enhance coagulation indirectly through proteolytic degradation of tissue factor pathway inhibitor, the major trigger protein in the onset of blood clotting.<sup>69</sup> Histones have also been shown to increase thrombin generation, causing platelet activation and coagulation.<sup>70</sup> In the experimental setting, administration of DNase1 to mice was found to promote disassembly of NETs with resultant suppression of DVT enlargement,<sup>71,72</sup> underscoring the importance of NETs in the pathogenesis of DVT.<sup>67</sup> The presence of NETs, according to the detection of citrullinated histone H3 positive cells, has also been described in human venous thrombi.<sup>73</sup>

### *NETs in lung disease*

NETs may also contribute to the pathogenesis and severity of several inflammatory lung conditions, including cystic fibrosis<sup>74</sup>, acute lung injury and acute respiratory distress syndrome<sup>75</sup>, severe asthma<sup>76</sup> and chronic obstructive pulmonary disease<sup>77</sup>.

#### Acute lung injury and acute respiratory distress syndrome

Acute lung injury (ALI) and acute respiratory distress syndrome (ARDS) represent a spectrum of lung diseases resulting from direct and indirect insults to the lung. These insults may be a result of infectious or sterile causes.<sup>78</sup> The disease process is characterised by a disruption of the endothelial-epithelial barriers, alveolar damage, pulmonary oedema and various degrees of respiratory failure.<sup>78</sup> ALI/ARDS is characterised by an influx of neutrophils into the pulmonary capillaries, with retention of hyperreactive neutrophils in the damaged vasculature.<sup>79</sup> NETs are able to contribute directly to the pathology of ALI/ARDS by inducing lung epithelial cell death.<sup>66</sup>

### *NETs in cystic fibrosis*

Cystic fibrosis (CF) is a lifelong inherited condition primarily affecting the lungs and digestive tract, with prevalence varying according to race, and seemingly more common in those of North European descent. CF patients develop chronic lung infections associated with airway obstruction mediated by viscous and insoluble mucus secretions.<sup>80</sup> In such patients, chronic bacterial colonisation of the airways develops, usually with the intransigent bacterial pathogens *S. aureus* and *P. aeruginosa*. Sputum viscosity is caused by extracellular DNA released from invading inflammatory cells, much of which is believed to originate from NETosis. This contention is supported by the observation that neutrophil elastase and MPO, which are found in high concentration in CF sputum, are bound to DNA, a key molecular signature of NETs.<sup>81</sup> If detached from NETs, neutrophil-derived proteolytic enzymes may also damage components of pulmonary connective tissue, especially elastin, compromising airway elasticity and function which may underpin the correlation between the magnitude of NET formation in the airways of CF patients and the degree of impairment of lung function.<sup>74,82</sup>

#### Pharmacological control of NETosis

Given the apparent involvement of excessive NETosis in the development of autoimmune and cardiovascular diseases, as well as in exacerbation of CF, pharmacological regulation of aberrant NETosis has definite therapeutic potential. This potential has yet to be realised, however, due in large part to the relatively recent discovery of NETosis, as well as the current limited insights into the diversity of the molecular mechanisms underpinning this process. Possible strategies include: (1) inhibition of the generation of pro-NETotic ROS, or, alternatively, neutralisation of ROS using oxidant-scavengers such as N-acetylcysteine or ascorbic acid;<sup>11</sup> (2) inhibitors of the activation of NF $\kappa$ B such as ascorbic acid and acetylsalicylic acid (aspirin), both of which have shown promise in experimental animal models of excessive NET formation<sup>11,83</sup>; and (3) inhibitors of PAD4 which are currently in pre-clinical development<sup>84</sup>.



With respect to alternative therapies, inhaled recombinant human DNase is widely used in the treatment of CF, primarily as a strategy to degrade neutrophil-derived DNA, a significant contributor to the viscosity of airway mucus.<sup>85</sup>

## Conclusions

NETs appear to increase the versatility and potency of the anti-infective armamentarium of neutrophils, as well as several other cell types of the innate immune system, possibly prolonging protective activity beyond cell death, thereby ensuring maximal utilisation of antimicrobial granule proteins. However, several important questions relating to the exact role of NETs in host defence remain incompletely understood. Remaining avenues for exploration include: (1) determining the biological relevance of the various types of NETosis; (2) unravelling the precise molecular and biochemical mechanisms underpinning these processes; and (3) characterising cooperative, beneficial interactions of NETosis with other cellular and humoral components of the innate and adaptive immune systems. As with other indiscriminate phagocyte-derived antimicrobial systems, such as the generation of tissue damaging and carcinogenic ROS, the extracellular release of nuclear material and enzymes such as elastase and MPO during NETosis presents the potential threat of development of autoimmune, cardiovascular and other disorders. Balancing NETosis in favour of host defence using pharmacological and other strategies represents an ongoing challenge.

## Authors' contributions

All the authors contributed equally to the compilation of the manuscript, as well as to the coordination of the final version.

## References

- Opitz B, Van Laak V, Eitel J, Suttrop N. Innate immune recognition in infectious and noninfectious diseases of the lung. *Am J Respir Crit Care Med*. 2010;181(12):1294–1309. <http://dx.doi.org/10.1164/rccm.200909-1427SO>
- Barber GN. Cytoplasmic DNA innate immune pathways. *Immunol Rev*. 2011;243(1):99–108. <http://dx.doi.org/10.1111/j.1600-065X.2011.01051.x>
- Brinkmann V, Reichard U, Goosmann C, Fauler B, Uhlemann Y, Weiss DS, et al. Neutrophil extracellular traps kill bacteria. *Science*. 2004;303(5663):1532–1535. <http://dx.doi.org/10.1126/science.1092385>
- Pillay J, Den Braber I, Vrískoop N, Kwast LM, De Boer RJ, Borghans JA, et al. In vivo labeling with <sup>2</sup>H<sub>2</sub>O reveals a human neutrophil lifespan of 5.4 days. *Blood*. 2010;116(4):625–627. <http://dx.doi.org/10.1182/blood-2010-01-259028>
- Robb CT, Dyrnynda EA, Gray RD, Rossi AG, Smith VJ. Invertebrate extracellular phagocyte traps show that chromatin is an ancient defence weapon. *Nat Commun*. 2014;5:4627. <http://dx.doi.org/10.1038/ncomms5627>
- Piłscek FH, Salina D, Poon KK, Fahey C, Yipp BG, Sibley CD, et al. A novel mechanism of rapid nuclear neutrophil extracellular trap formation in response to *Staphylococcus aureus*. *J Immunol*. 2010;185(12):7413–7425. <http://dx.doi.org/10.4049/jimmunol.1000675>
- Parker H, Dragunow M, Hampton MB, Kettle AJ, Winterbourn CC. Requirements for NADPH oxidase and myeloperoxidase in neutrophil extracellular trap formation differ depending on the stimulus. *J Leukoc Biol*. 2012;92(4):841–849. <http://dx.doi.org/10.1189/jlb.1211601>
- Yousefi S, Mihalache C, Kozłowski E, Schmid I, Simon HU. Viable neutrophils release mitochondrial DNA to form neutrophil extracellular traps. *Cell Death Differ*. 2009;16(11):1438–1444. <http://dx.doi.org/10.1038/cdd.2009.96>
- Fuchs TA, Abed U, Goosmann C, Hurwitz R, Schulze I, Wahn V, et al. Novel cell death program leads to neutrophil extracellular traps. *J Cell Biol*. 2007;176(2):231–241. <http://dx.doi.org/10.1083/jcb.200606027>
- Hakkim A, Fuchs TA, Martinez NE, Hess S, Prinz H, Zychlinsky A, et al. Activation of the Raf-MEK-ERK pathway is required for neutrophil extracellular trap formation. *Nat Chem Biol*. 2011;7(2):75–77. <http://dx.doi.org/10.1038/nchembio.496>
- Mohammed BM, Fisher BJ, Kraskauskas D, Farkas D, Brophy DF, Fowler AA, et al. Vitamin C: A novel regulator of neutrophil extracellular trap formation. *Nutrients*. 2013;5(8):3131–3151. <http://dx.doi.org/10.3390/nu5083131>
- Wang Y, Li M, Stadler S, Correll S, Li P, Wang D, et al. Histone hypercitullination mediates chromatin decondensation and neutrophil extracellular trap formation. *J Cell Biol*. 2009;184(2):205–213. <http://dx.doi.org/10.1083/jcb.200806072>
- Leshner M, Wang S, Lewis C, Zheng H, Chen XA, Santy L, et al. PAD4 mediated histone hypercitullination induces heterochromatin decondensation and chromatin unfolding to form neutrophil extracellular trap-like structures. *Front Immunol*. 2012;3, Art. #307, 11 pages. <http://dx.doi.org/10.3389/fimmu.2012.00307>
- Metzler KD, Goosmann C, Lubojemska A, Zychlinsky A, Papayannopoulos V. A myeloperoxidase-containing complex regulates neutrophil elastase release and actin dynamics during NETosis. *Cell Rep*. 2014;8(3):883–896. <http://dx.doi.org/10.1016/j.celrep.2014.06.044>
- Neumann A, Berends ET, Nerlich A, Molhoek EM, Gallo RL, Meerloo T, et al. The antimicrobial peptide LL-37 facilitates the formation of neutrophil extracellular traps. *Biochem J*. 2014;464(1):3–11. <http://dx.doi.org/10.1042/BJ20140778>
- Neumann A, Vollger L, Berends ET, Molhoek EM, Stapels DA, Midon M, et al. Novel role of the antimicrobial peptide LL-37 in the protection of neutrophil extracellular traps against degradation by bacterial nucleases. *J Innate Immun*. 2014;6(6):860–868. <http://dx.doi.org/10.1159/000363699>
- Clark SR, Ma AC, Tavener SA, McDonald B, Goodarzi Z, Kelly MM, et al. Platelet TLR4 activates neutrophil extracellular traps to ensnare bacteria in septic blood. *Nat Med*. 2007;13(4):463–469. <http://dx.doi.org/10.1038/nm1565>
- Yipp BG, Petri B, Salina D, Jenne CN, Scott BN, Zbytniuk LD, et al. Infection-induced NETosis is a dynamic process involving neutrophil multitasking in vivo. *Nat Med*. 2012;18(9):1386–1393. <http://dx.doi.org/10.1038/nm.2847>
- Byrd AS, O'Brien XM, Johnson CM, Lavigne LM, Reichner JS. An extracellular matrix-based mechanism of rapid neutrophil extracellular trap formation in response to *Candida albicans*. *J Immunol*. 2013;190(8):4136–4148. <http://dx.doi.org/10.4049/jimmunol.1202671>
- Yipp BG, Kubes P. NETosis: How vital is it? *Blood*. 2013;122(16):2784–2794. <http://dx.doi.org/10.1182/blood-2013-04-457671>
- Arai Y, Nishinaka Y, Arai T, Morita M, Mizuguchi K, Adachi S, et al. Uric acid induces NADPH oxidase-independent neutrophil extracellular trap formation. *Biochem Biophys Res Commun*. 2014;443(2):556–561. <http://dx.doi.org/10.1016/j.bbrc.2013.12.007>
- Itagaki K, Kaczmarek E, Lee YT, Tang IT, Isal B, Adibnia Y, et al. Mitochondrial DNA released by trauma induces neutrophil extracellular traps. *PLoS One*. 2015;10(3), e0120549, 10 pages. <http://dx.doi.org/10.1371/journal.pone.0120549>
- Hirsch JG. Bactericidal action of histone. *J Exp Med*. 1958;108(6):925–944. <http://dx.doi.org/10.1084/jem.108.6.925>
- Urban CF, Ermert D, Schmid M, Abu-Abed U, Goosmann C, Nacken W, et al. Neutrophil extracellular traps contain calprotectin, a cytosolic protein complex involved in host defense against *Candida albicans*. *PLoS Pathog*. 2009;5(10), e1000639, 18 pages. <http://dx.doi.org/10.1371/journal.ppat.1000639>
- Halverson TW, Wilton M, Poon KK, Petri B, Lewenza S. DNA is an antimicrobial component of neutrophil extracellular traps. *PLoS Pathog*. 2015;11(1), e1004593, 23 pages. <http://dx.doi.org/10.1371/journal.ppat.1004593>
- Rahman S, Gadjeva M. Does NETosis contribute to the bacterial pathoadaptation in cystic fibrosis? *Front Immunol*. 2014;5, Art. #378, 6 pages. <http://dx.doi.org/10.3389/fimmu.2014.00378>
- Young RL, Malcolm KC, Kret JE, Caceres SM, Poch KR, Nichols DP, et al. Neutrophil extracellular trap (NET)-mediated killing of *Pseudomonas aeruginosa*: Evidence of acquired resistance within the CF airway, independent of CFTR. *PLoS One*. 2011;6(9), e23637, 10 pages. <http://dx.doi.org/10.1371/journal.pone.0023637>
- Menten-Dedoyart C, Faccinnetto C, Golovchenko M, Dupiereux I, Van Lerberghe PB, Dubois S, et al. Neutrophil extracellular traps entrap and kill *Borrelia burgdorferi* sensu stricto spirochetes and are not affected by *Ixodes ricinus* tick saliva. *J Immunol*. 2012;189(11):5393–5401. <http://dx.doi.org/10.4049/jimmunol.1103771>

29. Riyapa D, Buddhisa S, Korbsrisate S, Cuccui J, Wren BW, Stevens MP, et al. Neutrophil extracellular traps exhibit antibacterial activity against *Burkholderia pseudomallei* and are influenced by bacterial and host factors. *Infect Immun*. 2012;80(11):3921–3929. <http://dx.doi.org/10.1128/IAI.00806-12>
30. Ramos-Kichik V, Mondragon-Flores R, Mondragon-Castelan M, Gonzalez-Pozos S, Muniz-Hernandez S, Rojas-Espinosa O, et al. Neutrophil extracellular traps are induced by *Mycobacterium tuberculosis*. *Tuberculosis (Edinb)*. 2009;89(1):29–37. <http://dx.doi.org/10.1016/j.tube.2008.09.009>
31. Branzk N, Lubojemska A, Hardison SE, Wang Q, Gutierrez MG, Brown GD, et al. Neutrophils sense microbe size and selectively release neutrophil extracellular traps in response to large pathogens. *Nat Immunol*. 2014;15(11):1017–1025. <http://dx.doi.org/10.1038/ni.2987>
32. Sumbly P, Barbian KD, Gardner DJ, Whitney AR, Welty DM, Long RD, et al. Extracellular deoxyribonuclease made by group A *Streptococcus* assists pathogenesis by enhancing evasion of the innate immune response. *Proc Natl Acad Sci USA*. 2005;102(5):1679–1684. <http://dx.doi.org/10.1073/pnas.0406641102>
33. Beiter K, Wartha F, Albiger B, Normark S, Zychlinsky A, Henriques-Normark B. An endonuclease allows *Streptococcus pneumoniae* to escape from neutrophil extracellular traps. *Curr Biol*. 2006;16(4):401–407. <http://dx.doi.org/10.1016/j.cub.2006.01.056>
34. Khatua B, Bhattacharya K, Mandal C. Sialoglycoproteins adsorbed by *Pseudomonas aeruginosa* facilitate their survival by impeding neutrophil extracellular trap through siglec-9. *J Leukoc Biol*. 2012;91(4):641–655. <http://dx.doi.org/10.1189/jlb.0511260>
35. Bruns S, Kniemeyer O, Hasenberg M, Amanianda V, Nietzsche S, Thywissen A, et al. Production of extracellular traps against *Aspergillus fumigatus* in vitro and in infected lung tissue is dependent on invading neutrophils and influenced by hydrophobin RodA. *PLoS Pathog*. 2010;6(4), e1000873, 18 pages. <http://dx.doi.org/10.1371/journal.ppat.1000873>
36. Wartha F, Beiter K, Albiger B, Fernebro J, Zychlinsky A, Normark S, et al. Capsule and D-alanylated lipoteichoic acids protect *Streptococcus pneumoniae* against neutrophil extracellular traps. *Cell Microbiol*. 2007;9(5):1162–1171. <http://dx.doi.org/10.1111/j.1462-5822.2006.00857.x>
37. Bianchi M, Niemiec MJ, Siler U, Urban CF, Reichenbach J. Restoration of anti-*Aspergillus* defense by neutrophil extracellular traps in human chronic granulomatous disease after gene therapy is calprotectin-dependent. *J Allergy Clin Immunol*. 2011;127(5):1243–1252e7. <http://dx.doi.org/10.1016/j.jaci.2011.01.021>
38. Shan Q, Dwyer M, Rahman S, Gadjeva M. Distinct susceptibilities of corneal *Pseudomonas aeruginosa* clinical isolates to neutrophil extracellular trap-mediated immunity. *Infect Immun*. 2014;82(10):4135–4143. <http://dx.doi.org/10.1128/IAI.02169-14>
39. Yoo DG, Floyd M, Winn M, Moskowitz SM, Rada B. NET formation induced by *Pseudomonas aeruginosa* cystic fibrosis isolates measured as release of myeloperoxidase-DNA and neutrophil elastase-DNA complexes. *Immunol Lett*. 2014;160(2):186–194. <http://dx.doi.org/10.1016/j.imlet.2014.03.003>
40. Thammarongsa V, Missiakas DM, Schneewind O. *Staphylococcus aureus* degrades neutrophil extracellular traps to promote immune cell death. *Science*. 2013;342(6160):863–866. <http://dx.doi.org/10.1126/science.1242255>
41. Lappann M, Danhof S, Guenther F, Olivares-Florez S, Mordhorst IL, Vogel U. In vitro resistance mechanisms of *Neisseria meningitidis* against neutrophil extracellular traps. *Mol Microbiol*. 2013;89(3):433–449. <http://dx.doi.org/10.1111/mmi.12288>
42. Juneau RA, Stevens JS, Apicella MA, Criss AK. A thermonuclease of *Neisseria gonorrhoeae* enhances bacterial escape from killing by neutrophil extracellular traps. *J Infect Dis*. 2015;212(2):316–324. <http://dx.doi.org/10.1093/infdis/jiv031>
43. Gunderson CW, Seifert HS. *Neisseria gonorrhoeae* elicits extracellular traps in primary neutrophil culture while suppressing the oxidative burst. *MBio*. 2015;6(1), e02452–14, 10 pages. <http://dx.doi.org/10.1128/mBio.02452-14>
44. Seper A, Hosseinzadeh A, Gorkiewicz G, Lichtenegger S, Roier S, Leitner DR, et al. *Vibrio cholerae* evades neutrophil extracellular traps by the activity of two extracellular nucleases. *PLoS Pathog*. 2013;9(9), e1003614, 15 pages. <http://dx.doi.org/10.1371/journal.ppat.1003614>
45. Guimaraes-Costa AB, DeSouza-Vieira TS, Paletta-Silva R, Freitas-Mesquita AL, Meyer-Fernandes JR, Saraiva EM. 3'-nucleotidase/nuclease activity allows *Leishmania* parasites to escape killing by neutrophil extracellular traps. *Infect Immun*. 2014;82(4):1732–1740. <http://dx.doi.org/10.1128/IAI.01232-13>
46. Abi Abdallah DS, Lin C, Ball CJ, King MR, Duhamel GE, Denkers EY. *Toxoplasma gondii* triggers release of human and mouse neutrophil extracellular traps. *Infect Immun*. 2012;80(2):768–777. <http://dx.doi.org/10.1128/IAI.05730-11>
47. Baker VS, Imade GE, Molta NB, Tawde P, Pam SD, Obadofin MO, et al. Cytokine-associated neutrophil extracellular traps and antinuclear antibodies in *Plasmodium falciparum* infected children under six years of age. *Malar J*. 2008;7, Art. #41, 12 pages. <http://dx.doi.org/10.1186/1475-2875-7-41>
48. Urban CF, Reichard U, Brinkmann V, Zychlinsky A. Neutrophil extracellular traps capture and kill *Candida albicans* yeast and hyphal forms. *Cell Microbiol*. 2006;8(4):668–676. <http://dx.doi.org/10.1111/j.1462-5822.2005.00659.x>
49. Kamoshida G, Kikuchi-Ueda T, Tansho-Nagakawa S, Nakano R, Nakano A, Kikuchi H, et al. *Acinetobacter baumannii* escape from neutrophil extracellular traps (NETs). *J Infect Chemother*. 2015;21(1):43–49. <http://dx.doi.org/10.1016/j.jiac.2014.08.032>
50. Rocha JD, Nascimento MT, Decote-Ricardo D, Corte-Real S, Morrot A, Heise N, et al. Capsular polysaccharides from *Cryptococcus neoformans* modulate production of neutrophil extracellular traps (NETs) by human neutrophils. *Sci Rep*. 2015;5, Art. #8008, 10 pages. <http://dx.doi.org/10.1038/srep08008>
51. Menegazzi R, Declava E, Dri P. Killing by neutrophil extracellular traps: Fact or folklore? *Blood*. 2012;119(5):1214–1216. <http://dx.doi.org/10.1182/blood-2011-07-364604>
52. Farrera C, Fadel B. Macrophage clearance of neutrophil extracellular traps is a silent process. *J Immunol*. 2013;191(5):2647–2656. <http://dx.doi.org/10.1093/jimmunol.1300436>
53. Wang H, Wang C, Zhao MH, Chen M. Neutrophil extracellular traps can activate alternative complement pathways. *Clin Exp Immunol*. 2015;181(3):518–527. <http://dx.doi.org/10.1111/cei.12654>
54. Saitoh T, Komano J, Saitoh Y, Misawa T, Takahama M, Kozaki T, et al. Neutrophil extracellular traps mediate a host defense response to human immunodeficiency virus-1. *Cell Host Microbe*. 2012;12(1):109–116. <http://dx.doi.org/10.1016/j.chom.2012.05.015>
55. Yost CC, Cody MJ, Harris ES, Thornton NL, McInturff AM, Martinez ML, et al. Impaired neutrophil extracellular trap (NET) formation: A novel innate immune deficiency of human neonates. *Blood*. 2009;113(25):6419–6427. <http://dx.doi.org/10.1182/blood-2008-07-171629>
56. Hazeldine J, Harris P, Chapple IL, Grant M, Greenwood H, Livesey A, et al. Impaired neutrophil extracellular trap formation: A novel defect in the innate immune system of aged individuals. *Aging Cell*. 2014;13(4):690–698. <http://dx.doi.org/10.1111/acel.12222>
57. Yu Y, Su K. Neutrophil extracellular traps and systemic lupus erythematosus. *J Clin Cell Immunol*. 2013;4(2), Art. #139, 7 pages. <http://dx.doi.org/10.4172/2155-9899.1000139>
58. Leffler J, Gullstrand B, Jonsen A, Nilsson JA, Martin M, Blom AM, et al. Degradation of neutrophil extracellular traps co-varies with disease activity in patients with systemic lupus erythematosus. *Arthritis Res Ther*. 2013;15(4), R84, 9 pages. <http://dx.doi.org/10.1186/ar4264>
59. Leffler J, Martin M, Gullstrand B, Tyden H, Lood C, Truedsson L, et al. Neutrophil extracellular traps that are not degraded in systemic lupus erythematosus activate complement exacerbating the disease. *J Immunol*. 2012;188(7):3522–3531. <http://dx.doi.org/10.4049/jimmunol.1102404>
60. Hakkim A, Furnrohr BG, Amann K, Laube B, Abed UA, Brinkmann V, et al. Impairment of neutrophil extracellular trap degradation is associated with lupus nephritis. *Proc Natl Acad Sci USA*. 2010;107(21):9813–9818. <http://dx.doi.org/10.1073/pnas.0909927107>
61. Carmona-Rivera C, Kaplan MJ. Low-density granulocytes: A distinct class of neutrophils in systemic autoimmunity. *Semin Immunopathol*. 2013;35(4):455–463. <http://dx.doi.org/10.1007/s00281-013-0375-7>
62. Khandpur R, Carmona-Rivera C, Vivekanandan-Giri A, Gizinski A, Yalavarthi S, Knight JS, et al. NETs are a source of citrullinated autoantigens and stimulate inflammatory responses in rheumatoid arthritis. *Sci Transl Med*. 2013;5(178):178ra40. <http://dx.doi.org/10.1126/scitranslmed.3005580>

63. Papadaki G, Choulaki C, Mitroulis I, Verginis P, Repa A, Raptopoulou A, et al. Enhanced release of neutrophil extracellular traps from peripheral blood neutrophils in patients with rheumatoid arthritis. *Ann Rheum Dis*. 2012;71(Suppl 1):A79. <http://dx.doi.org/10.1136/annrheumdis-2011-201238.17>
64. Zawrotniak M, Rapala-Kozik M. Neutrophil extracellular traps (NETs) – Formation and implications. *Acta Biochim Pol*. 2013;60(3):277–284.
65. Kessenbrock K, Krumbholz M, Schonermarck U, Back W, Gross WL, Werb Z, et al. Netting neutrophils in autoimmune small-vessel vasculitis. *Nat Med*. 2009;15(6):623–625. <http://dx.doi.org/10.1038/nm.1959>
66. Saffarzadeh M, Juenemann C, Queisser MA, Lochnit G, Barreto G, Galuska SP et al. Neutrophil extracellular traps directly induce epithelial and endothelial cell death: A predominant role of histones. *PLoS One*. 2012;7(2), e32366, 14 pages. <http://dx.doi.org/10.1371/journal.pone.0032366>
67. Fuchs TA, Brill A, Wagner DD. Neutrophil extracellular trap (NET) impact on deep vein thrombosis. *Arterioscler Thromb Vasc Biol*. 2012;32(8):1777–1783. <http://dx.doi.org/10.1161/ATVBAHA.111.242859>
68. Fuchs TA, Brill A, Duerschmied D, Schatzberg D, Monestier M, Myers DD Jr., et al. Extracellular DNA traps promote thrombosis. *Proc Natl Acad Sci USA*. 2010;107(36):15880–15885. <http://dx.doi.org/10.1073/pnas.1005743107>
69. Saffarzadeh M, Preissner KT. Fighting against the dark side of neutrophil extracellular traps in disease: Manoeuvres for host protection. *Curr Opin Hematol*. 2013;20(1):3–9. <http://dx.doi.org/10.1097/MOH.0b013e32835a0025>
70. Martinod K, Wagner DD. Thrombosis: Tangled up in NETs. *Blood*. 2014;123(18):2768–2776. <http://dx.doi.org/10.1182/blood-2013-10-463646>
71. Brill A, Fuchs TA, Savchenko AS, Thomas GM, Martinod K, De Meyer SF, et al. Neutrophil extracellular traps promote deep vein thrombosis in mice. *J Thromb Haemost*. 2012;10(1):136–144. <http://dx.doi.org/10.1111/j.1538-7836.2011.04544.x>
72. Von Bruhl ML, Stark K, Steinhart A, Chandraratne S, Konrad I, Lorenz M, et al. Monocytes, neutrophils, and platelets cooperate to initiate and propagate venous thrombosis in mice in vivo. *J Exp Med*. 2012;209(4):819–835. <http://dx.doi.org/10.1084/jem.20112322>
73. Savchenko AS, Martinod K, Seidman MA, Wong SL, Borissoff JI, Piazza G, et al. Neutrophil extracellular traps form predominantly during the organizing stage of human venous thromboembolism development. *J Thromb Haemost*. 2014;12(6):860–870. <http://dx.doi.org/10.1111/jth.12571>
74. Marcos V, Zhou Z, Yildirim AO, Bohla A, Hector A, Vitkov L, et al. CXCR2 mediates NADPH oxidase-independent neutrophil extracellular trap formation in cystic fibrosis airway inflammation. *Nat Med*. 2010;16(9):1018–1023. <http://dx.doi.org/10.1038/nm.2209>
75. Cheng OZ, Palaniyar N. NET balancing: A problem in inflammatory lung diseases. *Front Immunol*. 2013;4, Art. #1, 13 pages. <http://dx.doi.org/10.3389/fimmu.2013.00001>
76. Dworski R, Simon HU, Hoskins A, Yousefi S. Eosinophil and neutrophil extracellular DNA traps in human allergic asthmatic airways. *J Allergy Clin Immunol*. 2011;127(5):1260–1266. <http://dx.doi.org/10.1016/j.jaci.2010.12.1103>
77. Grabcanovic-Musija F, Obermayer A, Stoiber W, Krautgartner WD, Steinbacher P, Winterberg N, et al. Neutrophil extracellular trap (NET) formation characterises stable and exacerbated COPD and correlates with airflow limitation. *Respir Res*. 2015;16, Art. #59, 12 pages. <http://dx.doi.org/10.1186/s12931-015-0221-7>
78. Dushianthan A, Grocott MP, Postle AD, Cusack R. Acute respiratory distress syndrome and acute lung injury. *Postgrad Med J*. 2011;87(1031):612–622. <http://dx.doi.org/10.1136/pgmj.2011.118398>
79. Summers C, Singh NR, White JF, Mackenzie IM, Johnston A, Solanki C, et al. Pulmonary retention of primed neutrophils: A novel protective host response, which is impaired in the acute respiratory distress syndrome. *Thorax*. 2014;69(7):623–629. <http://dx.doi.org/10.1136/thoraxjnl-2013-204742>
80. Khatri IA, Bhaskar KR, Lamont JT, Sajjan SU, Ho CK, Forstner J. Effect of chondroitinase ABC on purulent sputum from cystic fibrosis and other patients. *Pediatr Res*. 2003;53(4):619–627. <http://dx.doi.org/10.1203/01.PDR.0000054780.11755.B9>
81. Papayannopoulos V, Staab D, Zychlinsky A. Neutrophil elastase enhances sputum solubilization in cystic fibrosis patients receiving DNase therapy. *PLoS One*. 2011;6(12), e28526, 7 pages. <http://dx.doi.org/10.1371/journal.pone.0028526>
82. Dwyer M, Shan Q, D'Ortona S, Maurer R, Mitchell R, Olesen H, et al. Cystic fibrosis sputum DNA has NETosis characteristics and neutrophil extracellular trap release is regulated by macrophage migration-inhibitory factor. *J Innate Immun*. 2014;6(6):765–779. <http://dx.doi.org/10.1159/000363242>
83. Laponni MJ, Carestia A, Landoni VI, Rivadeneyra L, Etulain J, Negrotto S, et al. Regulation of neutrophil extracellular trap formation by anti-inflammatory drugs. *J Pharmacol Exp Ther*. 2013;345(3):430–437. <http://dx.doi.org/10.1124/jpet.112.202879>
84. Bozdog M, Dreker T, Henry C, Tosco P, Vallaro M, Fruttero R, et al. Novel small molecule protein arginine deiminase 4 (PAD4) inhibitors. *Bioorg Med Chem Lett*. 2013;23(3):715–719. <http://dx.doi.org/10.1016/j.bmcl.2012.11.102>
85. Bakker EM, Volpi S, Salonini E, Van der Wiel-Kooij EC, Sintnicolaas CJ, Hop WC, et al. Improved treatment response to dornase alfa in cystic fibrosis patients using controlled inhalation. *Eur Respir J*. 2011;38(6):1328–1335. <http://dx.doi.org/10.1183/09031936.00006211>



# The botanical content in the South African curriculum: A barren desert or a thriving forest?

## AUTHOR:

Amelia L. Abrie<sup>1</sup>

## AFFILIATION:

<sup>1</sup>Department of Science, Mathematics and Technology Education, University of Pretoria, Pretoria, South Africa

## CORRESPONDENCE TO:

Amelia Abrie

## EMAIL:

Mia.Abrie@up.ac.za

## POSTAL ADDRESS:

Department of Science, Mathematics and Technology Education, Groenkloof Campus, University of Pretoria, Private Bag X20, Hatfield 0028, South Africa

## DATES:

Received: 30 Mar. 2015

Revised: 22 July 2015

Accepted: 06 Aug. 2015

## KEYWORDS:

plant sciences; botanical education; life sciences curriculum; Grades R–12 plant blindness

## HOW TO CITE:

Abrie AL. The botanical content in the South African curriculum: A barren desert or a thriving forest? *S Afr J Sci.* 2016;112(1/2), Art. #2015-0127, 7 pages. <http://dx.doi.org/10.17159/sajs.2016/20150127>

Botanists who are interested in education have often expressed their dismay at how plant sciences are neglected in Biology curricula, despite the important roles that plants play. While botanists in several overseas countries have studied the ways in which plant sciences are represented in curricula, no research has been done on how botany is neglected in the South African curriculum. Currently, the South African curriculum is known as the Curriculum and Assessment Policy Statements (CAPS) for Grades R–12. In this study, a comparison was made among the content that is generally taught in introductory plant sciences courses, the American Society of Plant Biologists' principles for plant biology education and the relevant CAPS documents. The time spent on plant, animal or human-focused content was established and compared at both phase and grade level. It was found that while the curriculum addresses all the major concepts in the plant sciences, very little time was being allocated to exclusively plant-focused content as compared to animal and human-focused content. This neglect was particularly prevalent in the Foundation Phase. The way in which the content is structured and presented in the curriculum may in all likelihood not be sufficient to provide a strong knowledge and skills foundation in the plant sciences, nor will it encourage the development of positive values towards plants. While consensus regarding the content of a curriculum will be difficult to achieve, awareness of potential gaps in the curriculum should be brought to the attention of the botanical and educational communities.

## Introduction

Wilkins<sup>1</sup> called plants the 'most important, least understood and most taken-for-granted of all living things' and botanists interested in education have been concerned about the neglect of plants in Biology curricula for a long time.<sup>2</sup> Indeed, Wandersee and Schussler<sup>3,4</sup> went so far as to develop the term 'plant blindness' to describe the lack of awareness of plants and neglect of plants in education. It should be noted too that there seems to be a disconnect between plant neglect in schools and the general population's interest in plants.<sup>5</sup> According to the US Census Bureau,<sup>6</sup> 68% of US households either bought plant and gardening-related products or engaged in gardening related activities in 2010 and according to a Harris poll<sup>7</sup>, 7% of US adults list gardening as a favourite leisure activity.

In the United States of America, botany experienced a 'golden age' in the early 1900s, but declined in popularity as disagreements about course content and pedagogy, as well as irrelevant content and a lack of qualified teachers caused a shift towards a single Biology subject.<sup>2</sup> Le Grange's<sup>8</sup> description of the history of Biology in South Africa confirms that similarly, at least in the Transvaal, Botany was the dominant subject in the life sciences and was only replaced by Biology as a school subject in 1935. Even though a strong emphasis on Botany remained for a considerable period, over time topics such as animal and human physiology and molecular biology were included in the subject, with less and less focus on plant sciences.

In South Africa, curriculum changes post-1994 included a move towards outcomes-based curricula.<sup>8,9</sup> In the process, Biology was renamed and became the Life Sciences. The most recent change has been the introduction of the Curriculum and Assessment Policy Statements (CAPS).<sup>10-13</sup> CAPS represents, once again, a performance-based syllabus type curriculum.<sup>14</sup> In this study, only the Life Sciences related CAPS documents were analysed.

These changes, coupled with the worldwide trend towards neglecting the plant sciences in Biology curricula, prompted this investigation into the consequences of the curriculum changes for the botanical sciences; specifically whether the plant sciences component of the curriculum has been weakened, or whether the study of plants has benefitted from the changes. An attempt was made to find a way in which to answer the question posed by Uno<sup>5</sup>: 'What and how should students learn about plants?' Subsequently, the plant-focused content in the relevant CAPS documents (Grades R–12) was assessed and compared with the prevalence of zoological and human-focused content.

## Literature

### *Why should students learn about plants?*

It has been shown in many studies that learners dislike plant sciences as a subject.<sup>2,4,5,15</sup> However, this does not mean that the plant sciences should be neglected in school curricula as we should teach what it is important for learners to know, not what they prefer to know.<sup>16</sup> Plants provide not only aesthetic pleasure, but many products that humans have exploited for millennia, such as food, clothing, construction materials, fuel, paper, industrial chemicals, alcohol, essential oils and medicines. These uses of plants should be enough reason to study plant sciences. In addition, many of the biggest challenges facing 21st century society are botanically based. These include deforestation, global warming, food security, the anthropocentric extinction of species, erosion, the influence of invasive species and the discovery of new plant-derived pharmaceuticals that may help in the fight against disease. Furthermore, plants are extensively used in biotechnology and are also often used to study fundamental life processes. Without a basic understanding of the structure, functioning, diversity and ecology



of plants, many of these benefits would be difficult to maintain. The US-based National Research Council<sup>17</sup> has encapsulated these views in a strong statement:

*Modern civilization rests on the successful and sustained cultivation of plants and on the wise use of the biologic and physical resource base on which their cultivation depends. Our knowledge about the world around us is incomplete if we do not include plants in our discoveries, and it is distorted if we do not place sufficient emphasis on plant life.*

### **Recontextualisation of the parent discipline to form the curriculum**

Bernstein<sup>18</sup> described the flow of information from the knowledge that is produced in the academic discipline in which it originated, to the form that the knowledge takes in the educational arena, including the school curriculum, and the classroom where the knowledge is reproduced as a process of recontextualisation. Although school science can be expected to differ from the original scientific discipline, it is reasonable to expect a strong resemblance with the parent discipline, even if it is, at best, a simplification. The knowledge that is incorporated in the school curriculum is described by Taber<sup>19</sup> as a curricular model of the science itself. The development of curricular models is an endeavour to find the optimum level of simplification – a level at which the content is simple enough for learners to understand and that also provides an adequate reflection of the scientific understanding of the parent discipline on which it is based. This refers not only to the depth of the content, but also to the choice of what should be included in and excluded from the curriculum.

Bernstein<sup>20</sup> classified those individuals who are involved in curriculum construction as belonging to the 'official recontextualising field' and he included those who recontextualise the curriculum to train teachers, write textbooks or conduct research on the curriculum in the 'pedagogic reconceptualising field'. It should be remembered that the construction of a curriculum is a selective process performed by those in the 'official recontextualising field' and as such, will be influenced and moulded by these individuals who have their own backgrounds, agendas and biases, even when the curriculum reflects the parent discipline to a reasonable degree. Kind and Taber<sup>21</sup> describe school science as politically constructed and assert that it cannot be a neutral reflection of professional science. Additionally, those who play a part in the 'pedagogical recontextualising field' will bring a further recontextualisation to the subject that is now encapsulated in the curriculum, further changing the content that will eventually reach the classroom. In this study, only the recontextualisation that occurs between the parent subject and the curriculum was considered.

The recontextualised curriculum must, furthermore, show curricular coherence if the logical structure of the parent discipline is to be made visible to students.<sup>22</sup> Curricular coherence implies that the subject matter is articulated over time as a sequence of topics and performances that reflect the logical and hierarchical nature of the subject. The content must show progression, in terms of depth and rigour, with connections and coordination between topics made apparent. No literature could be found that explores the curricular coherence in the current Life Sciences curriculum and whether there is a coherent structure to the content related to plants.

### **What should students learn about plants?**

One of the primary questions faced by those attempting to study the recontextualisation of the Life Sciences into a comprehensive, yet simplified curriculum concerns the content. It is not easy to decide what learners should know about plants and this is one of the challenges that curriculum designers face. Uno<sup>5</sup> suggested that new and existing knowledge should be structured around the major concepts and principles of the discipline and cautioned that it would be difficult to agree on a consensus list of facts or terms that students would have to know. As a result, he did not provide clarity on what these major concepts and principles might be. The American Society of Plant Biologists (ASPB)<sup>23</sup>

developed a list of basic plant biology principles for science education at the K–12 level (see Table 1). This list of concepts essentially refers to content for science education that can be integrated with biology at different educational levels.

Khodor, Halme and Walker<sup>24</sup> developed a biology concept framework as one possible way to organise the large number of concepts that are covered by a typical undergraduate Biology curriculum. This framework 'is hierarchical, places details in context, nests related concepts, and articulates concepts that are inherently obvious to experts, but often difficult for novices to grasp'. Unfortunately, no similar concept framework has been developed for the plant sciences. One factor that complicates the development of such a framework is the inevitable overlap between the concepts that are essential for the understanding of the plant sciences and the concepts that are common to other biological sciences, especially the zoological sciences and human physiology.

As it was not the purpose of this study to develop a complete concept framework for the plant sciences, similar to the biology concept framework, another strategy had to be followed. It was decided to follow the example of UMALUSI<sup>14</sup> and develop a list of concepts or topics that reflects typical botanical content from tertiary level textbooks that are specifically written by botanists to be used in undergraduate plant sciences courses. While this is not a comprehensive, definitive list of concepts, it provides a useful tool with which to analyse the plant sciences content in the CAPS curriculum.

### **Method**

Four undergraduate textbooks were used to develop a list of common topics that could be described as representative of a typical undergraduate plant sciences course. These textbooks are generally available from publishing houses in South Africa and all have been previously prescribed for plant sciences related modules by the author. The textbooks all originate from the United States of America and are general texts that are used in undergraduate plant sciences courses. The number and variety of unique plant-related course offerings in different South African universities does not allow the author to draw conclusions about the extent to which each book is currently prescribed in South Africa. The latest editions that are available were used. Most of the information was gained from the index pages, although in some cases, some clarification had to be sought from the text where the headings were not sufficiently descriptive. Textbooks by the following authors were used: Mauseth<sup>25</sup>, Evert and Eichhorn<sup>26</sup>, Stern et al.<sup>27</sup> and Uno et al.<sup>28</sup> The ASPBs<sup>23</sup> list of principles of plant biology was mapped against this topic list.

The study used document analysis to determine what botanical content is included in those CAPS subjects that either include science content or life sciences in particular. All content related to plants or plant sciences was identified, including content relevant to other living organisms. In order to find out whether the topics represented the main botanical ideas, they were compared with the list of common plant science topics and the ASPB list. It should be noted that the main focus in this article is on the *intended* curriculum as it is encapsulated in the various curriculum documents and not on the *enacted* curriculum. The following Curriculum and Assessment Policy Statements (CAPS documents) were included:

CAPS Life Skills Foundation Phase Grades R–3<sup>10</sup>

CAPS Natural Sciences and Technology Intermediate Phase Grades 4–6<sup>11</sup>

CAPS Natural Sciences Senior Phase Grades 7–9<sup>12</sup>

CAPS Life Sciences Further Education and Training Phase Grades 10–12<sup>13</sup>

The CAPS documents were examined to compare the amount of plant-focused content with the animal- or human-focused content. The number of hours that the curriculum specifies should be dedicated to each topic was classified either as being generally applicable to all or most living organisms, as being applicable to plant-focused content

**Table 1:** The principles of plant biology developed by the American Society of Plant Biologists.<sup>23</sup>

1	Plants contain the same biological processes and biochemistry as microbes and animals. However, plants are unique in that they have the ability to use energy from sunlight along with other chemical elements for growth. This process of photosynthesis provides the world's supply of food and energy.
2	Plants require certain inorganic elements for growth and play an essential role in the circulation of these nutrients within the biosphere.
3	Land plants evolved from ocean-dwelling, algae-like ancestors, and plants have played a role in the evolution of life, including the addition of oxygen and ozone to the atmosphere.
4	Reproduction in flowering plants takes place sexually, resulting in the production of a seed. Reproduction can also occur via asexual propagation.
5	Plants, like animals and many microbes, respire and utilise energy to grow and reproduce.
6	Cell walls provide structural support for the plant and also provide fibres and building materials for humans, insects, birds and many other organisms.
7	Plants exhibit diversity in size and shape ranging from single cells to gigantic trees.
8	Plants are a primary source of fibre, medicines, and countless other important products in everyday use.
9	Plants, like animals, are subject to injury and death due to infectious diseases caused by microorganisms. Plants have unique ways to defend themselves against pests and diseases.
10	Water is the major molecule present in plant cells and organs. In addition to an essential role in plant structure, development, and growth, water can be important for the internal circulation of organic molecules and salts.
11	Plant growth and development are under the control of hormones and can be affected by external signals such as light, gravity, touch, or environmental stresses.
12	Plants live and adapt to a wide variety of environments. Plants provide diverse habitats for birds, beneficial insects, and other wildlife in ecosystems.

exclusively, as being applicable to animal-focused content exclusively, or as being applicable to human-focused content (and thus often applicable to animals too). Analysis along these lines was conducted for the four phases of schooling and for the individual grades.

## Results

The list of topics and concepts that the four botany textbooks had in common is shown in the first column of Table 2. The textbooks were very similar in their approaches to the plant sciences and contrary to Uno's<sup>5</sup> prediction that it would be difficult to agree on a consensus list of topics, it would appear that there is broad consensus about the content that should be covered in a typical undergraduate plant sciences course. The amount of detail and the depth to which each topic is covered varies between the textbooks, but the major concepts that students are required to study are clearly identifiable. The content in these textbooks is considerably more advanced than would be taught at school level, but still provides an indication of the breadth and content of the subject. A comparison between the textbook topic list and the ASPB<sup>23</sup> list of principles of plant biology can be drawn from the first and the second columns of Table 2. It is clear that the ASPB list of principles is not as extensive as the textbook topic list, but it still covers all the major ideas in the plant sciences. The grades in which these topics are included are shown in the third column of Table 2.

A comparison between the general, animal, human and plant-focused content showed that exclusively plant-focused topics were the least prominent in the South African curriculum.

### *Foundation Phase Life Sciences – the early years*

The Curriculum and Assessment Policy Statement for Life Skills - Foundation Phase<sup>10</sup> sets the zoocentric tone for the school curriculum. In the Foundation Phase, the strand 'Beginning knowledge and Personal and Social Well-being' that forms part of the life skills subject includes topics that introduce learners to content that will form the foundation for natural sciences topics in Grades 4–9 and life sciences topics in Grades 10–12. This part of the curriculum is understandably dominated by human-focused topics (e.g. 'my body' and 'the senses'), that are of course also

valid for many animals, especially mammals. The curriculum includes very little plant-focused content, while animal-focused content is much more prevalent. The few references to plants are mostly associated with how plants provide humans (and animals) with food, shade or shelter. Plants are not shown as interesting organisms in their own right. The time spent on plant-related content in Grades R–3 is approximately 11 h. It is not possible to provide a precise number for the hours required to teach plant-focused content, as this content is frequently integrated into more general topics such as 'the seasons'. Furthermore, some of the curriculum content may or may not include plant-focused content, depending on the teachers' interpretation of the content. In contrast, animal-related topics feature regularly in the life skills curriculum and occupy approximately 57 h of teaching and learning time.

The only exclusively plant-focused topic in Grade 1, the 'basic structure of plants and what they need to grow, is covered in 4 h. In Grade 2, learners again study the effects of the seasons on organisms, this time with reference to the sowing, growing and harvesting of plants. No exclusively plant-focused content is included in Grade 3, unless the sections on healthy eating and the processing of products that we get from plants are regarded as plant focused. In contrast, 15 h are devoted to insects and animal life cycles, more than all the plant-focused content for Grades R–3 combined.

In the Foundation Phase, routine, free play or structured activities are recognised. Only one of the free play activities, gardening, is plant focused. Plants do not feature in the resource list for life skills<sup>10</sup>(p. 13–14), but recommended resources do include a selection of fruit and vegetables (p. 20) and a suggestion for a nature corner/discovery table with plants or pictures of plants for the Grade 1 topic of plants and seeds (p. 33).

To appreciate the contrast between the plant and animal sciences, a comparison shows that Grade R learners spend 2 h each on birds, reptiles, dinosaurs, wild animals in general and finding out about one wild animal in particular, while the teacher may or may not include plants in topics on how nature is affected by the seasons. The only additional topic in Grade R that is plant related is the section on fruit and vegetables,

**Table 2:** List of the topics in the Curriculum and Assessment Policy Statements (CAPS) curricula for Grades R–12<sup>10-13</sup> that are included in botany textbooks, including those by Mauseth<sup>25</sup>, Evert and Eichhorn<sup>26</sup>, Stern et al.<sup>27</sup> and Uno et al.<sup>28</sup> The American Society of Plant Biologists (ASPB)<sup>23</sup> principles listed in Table 1 that correspond to these topics are shown in the second column. It should be noted that not all the ASPB topics are covered in detail.

Topics generally covered in plant sciences textbooks	ASPB principle number	Grade/s in which the content is included in the CAPS curriculum
Introduction		
Introduction to plants		1, 4
The Nature of Science*		10
Plant-human interactions	6, 7	1, 2, 6, 12
Life at the molecular and cellular level		
Chemistry of life*	1	10
Cell structure*	6	9, 10
Cell division, mitosis*		10
Genetics and heredity		
Chemistry of heredity, DNA*	1	12
Genetics and patterns of inheritance*		12
Recombinant DNA technology and biotechnology*	8	10, 12
Meiosis*		12
Plant structure		
Plant tissues and organs	6	R, 1, 4, 9, 10
Growth and development	7	4, 5
Life processes in plants		
Reproduction: flowers, fruit and seeds	4	R, 1, 7, 11
Mineral nutrition and soil	2	1, 2, 10
Water relations and transport of nutrients	2, 10	10
Energy metabolism: Photosynthesis and respiration	1, 5	5, 6, 8, 11
Hormones – responding to the environment	11	12
Classification and biodiversity		
Classification and systematics*		7, 10
Biodiversity*		7, 10
Viruses*		8, 11
Prokaryotes: Bacteria and archaea*		7, 8, 11
Fungi*		8, 11
Algae*	7	11
Non-vascular plants: mosses	7	11
Vascular, seedless plants: ferns	7	7, 11
Vascular seed plants: Gymnosperms and angiosperms	7	7, 11
Ecology – life at the population, community and ecosystem level		
Introduction to ecology*		8
Ecosystems and environmental interactions*	2, 12	R, 1, 4, 5, 6, 8, 10, 11
Food chains and food webs*		5, 6, 8, 10
Population ecology*		11
Biomes	12	10
Biosphere*	12	7, 10
Conservation*		8
Human impact on the environment*		11
Evolution		
Evolutionary concepts*		7,
History of ideas about origins*		12
History of life on earth*	3	10
Fossils*		10, 12

\* Note that not all the topics in these textbooks are exclusively focused on plant sciences. These topics are applicable to other organisms too.

where the focus is placed on types, tastes, textures, colours and shapes, with a single reference to where fruit and vegetables come from.

### *Intermediate Phase: Grades 4–6*

In the Intermediate Phase, science topics are addressed in the subject Natural Sciences and Technology which requires 3.5 h of instructional time per week. The subject is divided into four knowledge strands, namely Life and Living, Matter and Materials, Energy and Change, and Planet Earth and Beyond.<sup>11(p.10)</sup> Life sciences are limited to the Life and Living strand which is taught only in the first term of each year (January to the end of March) – a total of 45 h each year.

Although the difference between exclusively plant-focused topics and animal-focused topics is smaller than in the Foundation Phase, there is still a preponderance of animal content in the Intermediate Phase. Many of the topics refer to plants and animals equally or at least do not exclude plants, for example: 'living and non-living things', 'food chains' and 'ecosystems and food webs'. These general topics require approximately 26,25 h of class time. It is left to the discretion of the teacher – and possibly to the writers of textbooks – how much of this content will be focused on animals instead of plants. Human Nutrition is taught for nearly 20 h. Exclusively animal-focused content (e.g. 'habitats of animals', 'structures for animal shelters' and 'animal skeletons') requires 33,75 h of teaching time. Exclusively plant-focused content (e.g. 'what plants need to grow' and 'photosynthesis'), occupies approximately 24,25 h.

The plant-focused content in this phase is somewhat repetitive and limited to dry, basic facts. The reference to photosynthesis in Grade 5 leads to the role of plants as producers (food) and in Grade 6, the same topic again leads to a discussion of food and oxygen and, although there is some progression, the content can easily be interpreted as repetitive. Plants are presented as mere providers of food and oxygen to be used by other organisms.

### *Senior Phase: Grades 7–9*

In the Senior Phase, the life sciences form part of the subject Natural Sciences. This subject includes the same four knowledge strands that are used in the Intermediate Phase<sup>12(p. 9)</sup>. The Grade 7, 8 and 9 Life and Living content is taught for 3 h per week during the first nine weeks of the school year only. Many of the topics, including biosphere, biodiversity, sexual reproduction, variation (Grade 7) and interactions and interdependence within the environment (Grade 8) appear to be focused on both plants and animals (including humans). The curriculum does not provide insight into the extent to which teachers will use this content to include plants and not only animals. A re-introduction to photosynthesis and respiration, as well as to micro-organisms, appears in Grade 8. The introduction to cells in Grade 9 requires 6 h of teaching time, with the remaining 21 h devoted to systems of the human body. In this phase, only 11,25 of the 81 h of teaching time is devoted to content that is specifically plant science focused.

The Grade 7 curriculum provides one example of the repetition of some of the plant-focused activities. Learners are required to grow plants such as beans or maize to determine the requirements for growth. This is the third time that learners have to perform this activity in school, having done so in Grade 1 (beans or lentils) and Grade 4 (bean seeds).

### *Further Education and Training: Grades 10–12*

In this final phase of schooling, Life Sciences is taught as a subject on its own and is divided into the following strands: Life at the Molecular, Cellular and Tissue level; Life Processes in Plants and Animals; Environmental Studies and Diversity, Change and Continuity.<sup>1</sup> A significant number of the topics can not be separated as either human, animal or plant focused, but are relevant to living organisms in general. Examples of such content include: how science works, the chemistry of life, DNA, cells, cell division, genetics and inheritance, biosphere to ecosystems, biodiversity and classification, history of life on earth, respiration, gaseous exchange, population ecology, human impact on the environment and evolution by natural selection (not listed in sequence).

These topics require 187 h of class time. The extent to which plants will be used as examples when these topics are taught in the enacted curriculum is not known, but previous research has shown that teachers prefer to use animal examples.<sup>3,4,16</sup> Exclusively human-focused topics (e.g. support and transport systems, human gas exchange, excretion, human endocrine system, and homeostasis in humans) are specified for approximately 76 h of teaching time. Animal-focused and plant-focused topics require approximately 52 h and 46 h respectively.<sup>13</sup>

## **Discussion**

It can be argued that the frequent changes in the policy and content of the curriculum have not provided sufficient opportunity for academic analysis of each reiteration. It is encouraging to see that the 4 CAPS documents that include the scientific content of the school curriculum give some attention to all the topics that are usually covered in plant sciences courses (see Table 2). Uno<sup>5</sup> suggests that new and existing knowledge should be structured around the major concepts and principles of the discipline for understanding to develop. However, during the analysis, it became clear that many of the topics received only cursory attention, whereas other topics, such as photosynthesis and ecosystems and environment interactions, were included in several grades and were covered in more depth, thus showing progression. The logic behind the sequencing of topics could not be discerned and connections and coordination between topics were not made apparent, thus failing to achieve curricular coherence as defined by Schmidt et al.<sup>22</sup>

Zoocentrism, anthropocentrism and plant blindness could be clearly identified from the data. The curriculum analysis provides an overview of the extent to which plant-focused content is overlooked in favour of animal and human-focused content in the life sciences related South African curriculum documents. Although the time that is currently devoted to the plant sciences is clearly more limited, it will be difficult, if not impossible to develop consensus in both the official recontextualising field and the pedagogic recontextualising field about the amount of time that should be devoted to different topics in the life sciences. However, the current lack of debate around the lack of representation of the plant sciences does not do justice to the potential consequences of such neglect.

The neglect is especially noticeable in the Foundation Phase where very little attention is given to plants in Grades R–2, with no exclusively plant-focused content in Grade 3. Children build their understanding of the biological world through their interactions with both the natural world and through exposure to school content and activities. Tunncliffe and Ueckert<sup>29</sup> recognise that very little is known about the development of botanical concepts and skills in the early years of schooling, as most of the research in biology education focuses on secondary school biology. The infrequent focus on plants and the way that content is distributed in the Foundation, Intermediate and Senior Phases will, in all likelihood, not be sufficient to provide a strong knowledge or skills foundation in the plant sciences. In a recent study, Foundation Phase learners in Gauteng even indicated that they do not think that plants are alive, but they do think that water is alive (Naudé F 2015, oral communication, July 21). According to Uno<sup>5</sup>, understanding is hampered when learners are exposed to disconnected facts, breadth rather than depth of coverage and a requirement for recall of facts. The long gap between one year and the start of the next year when learners will once more study the life sciences may not encourage continuity in the development of a coherent understanding of the knowledge and skills associated with the subject in general and plant sciences specifically.

The lack of plant science content will probably hamper the development of positive values towards plants, but there is very little research concerning the development of such values in the early years. The development of a relationship with the environment, including positive values and attitudes, is included in the explanation of the Foundation Phase Life Skills<sup>10(p. 9–10)</sup> subject, but it is difficult to see how the limited opportunities that learners have to learn to appreciate plants will encourage the development of this relationship. It is important that learners are shown the relevance of the content that they learn in order to stimulate their interest in the subject.<sup>5</sup> This means that learners should be shown, not only how useful plants are to humans, but also that it is



important to appreciate plants as the interesting organisms they are. It is even possible, given the status quo, that learners will develop the idea that plants are not worthy of being studied.

Several authors have offered suggestions on how learners' plant blindness may be averted. De Beer and Van Wyk<sup>30</sup> suggested that teachers can stimulate learners' interest in plants and encourage learners to consider a career in the plant sciences by introducing ethnobotany into the classroom. They provide several examples of the judicious use of South African ethnobotanical content that will not only contribute to the inclusion of indigenous knowledge into the curriculum, but that will increase the relevance of plants to learners while avoiding repetitive, dry factual content. Both Hershey<sup>16</sup> and Strgar<sup>31</sup> note that appreciation of plants may benefit from the guidance and enthusiasm of a teacher or mentor. Tunniff and Uecker<sup>29</sup> suggest that children must be assisted to look with meaning and accuracy, in contrast with merely seeing. In the case of plants, this is especially important, as research has shown that 'children may not see the trees'<sup>32</sup> and, unlike animals that draw attention to themselves, 'plants need to have attention drawn to them'.<sup>15</sup> Balas and Momsen<sup>33</sup> demonstrated that there is a real difference in the way that plants and animals are perceived and that plant blindness is, in part, a physiological phenomenon. However, as the plant sciences content in curricula, including CAPS, has decreased over time, the lack of plant-related content, coupled with decreasing appreciation of plants, may have become a self-perpetuating phenomenon leading to teachers who themselves lack the appropriate values to encourage appreciation in learners.

It is not known whether the use of the suggested activities, equipment and resources that complement the teaching of the CAPS curricula is interpreted as a compulsory part of the intended curriculum, or whether teachers are allowed flexibility to choose different activities. Several of the suggested activities may be beneficial to the development of a love of plants. Furthermore, if they are conducted in a structured and systematic way, knowledge, skills and understanding of plants can be enhanced while the development of misconceptions can be minimised. Pedagogical approaches to learning and teaching that emphasise critical thinking and process skills, combined with an understanding of concepts, will contribute to the development of botanically literate students and citizens.<sup>5</sup> It is essential that teachers understand that seemingly simple activities, such as observing plants, learning the names of a few plants, labelling, describing and comparing the parts of different plants, or germinating seeds, may counteract some of the gaps in basic botanical knowledge.<sup>29</sup> These activities may lead to the development of knowledge and skills that promote inquiry in more complex topics, such as the community structure and may also allow the development of a closer bond between the learner and nature.<sup>34,35</sup> Whether the activities described in the CAPS documents are indeed sufficient to allow learners to develop knowledge, skills and a positive view of plants remains to be investigated in the South African context.

## Conclusion

This purpose of this paper was not to argue for a specific 'ideal' curriculum, but to explore the existing curriculum in terms of the botanical content contained in the CAPS documents. In the CAPS for both Foundation and Intermediate Phases, very little attention is given to plants. In the Senior and FET Phases, the situation is not much better. The ASPB<sup>23</sup> cautions that childhood is the time when persistent inaccurate ideas about plants can be established and they warn that a basic understanding of how plants live and grow is essential for the understanding of the many environmental challenges that humanity faces in the 21st century.

An extensive search of the South African literature did not provide any articles that highlighted the plight of the plant sciences in school curricula, although a growing literature base concerning this phenomenon is developing in Europe and the USA. The lack of plant-focused content, especially in the early years of schooling should be of great concern, not only to educators, but to botanists who are concerned with the way that plants are viewed in the community, and even to parents who may

recognise that a biased curriculum will not provide their children with a balanced understanding of the life sciences.

In some countries such as the United States of America, the United Kingdom and Sweden, there has been some interest in the way that botany is portrayed in curricula. The concerns regarding the lack of botanical instruction raised by the ASPB<sup>23</sup> and other authors referred to above seems not to have reached the botanical community in South Africa. As far as could be determined in an extensive literature survey, there has been no comment from the botanical community in South Africa regarding the lack of focus on the plant sciences in the curriculum documents. As Hoekstra<sup>36</sup> suggests, 'botanists work very hard to make their science second-rate in the eyes of the public' – a sentiment shared by Hershey<sup>16</sup> who declared: 'Plant blindness: "We have met the enemy and He is Us"'.<sup>16</sup>

## References

1. Wilkens M. *Plantwatching: How plants remember, tell time, form relationships and more*. New York: Facts On File; 1988.
2. Hershey DR. An historical perspective on problems in botany teaching. *Am Biol Teach*. 1996;58:340–347. <http://dx.doi.org/10.2307/4450174>
3. Wandersee JH, Schussler EE. Preventing plant blindness. *Am Biol Teach*. 1999;61:84–86. <http://dx.doi.org/10.2307/4450624>
4. Wandersee JH, Schussler EE. Toward a theory of plant blindness. *Plant Sci Bull*. 2001;47(1):2–9.
5. Uno GE. Botanical literacy: What and how should students learn about plants? *Am J Bot*. 2009;96(10):1753–1759. <http://dx.doi.org/10.3732/ajb.0900025>
6. US Census Bureau. Statistical abstract of the United States: 2012. Section 26 arts, recreation and travel: Recreation and leisure Activities [document on the Internet]. c2012 [cited 2015 Feb 25]. Available from: <https://www.census.gov/prod/2011pubs/12statab/arts.pdf>
7. The Harris Poll. We work before we play: Kids in the household? Then, what is leisure time? [document on the Internet]. c2013 [cited 2015 Mar 15]. Available from: [http://www.harrisinteractive.com/vault/Harris%20Poll%2093%20-%20Work%20and%20Leisure\\_12.9.2013.pdf](http://www.harrisinteractive.com/vault/Harris%20Poll%2093%20-%20Work%20and%20Leisure_12.9.2013.pdf)
8. Le Grange L. The history of biology as a school subject and developments in the subject in contemporary South Africa. *SARE*. 2008;14(3):89–105.
9. Le Grange L. Veranderinge in skoolbiologie in Suid-Afrika ná apartheid. [Changes in school biology in South Africa after 'apartheid'.] *S-Afr Tydskr Natuurwet Tegnol*. 2012;31(1), Art. #385, 8 pages. Afrikaans. Available from: <http://www.satnt.ac.za/index.php/satnt/article/view/385/534> / <http://dx.doi.org/10.4102/satnt.v31i1.385>
10. Department of Basic Education. Curriculum and Assessment Policy Statement (CAPS): Life Skills. Foundation Phase grades R to 3. Pretoria: Government Printing Works; 2011.
11. Department of Basic Education. Curriculum and Assessment Policy Statement (CAPS): Natural Sciences and Technology. Intermediate Phase grades 4–6. Pretoria: Government Printing Works; 2011.
12. Department of Basic Education. Curriculum and Assessment Policy Statement (CAPS): Natural Sciences. Senior Phase grades 7–9. Pretoria: Government Printing Works; 2011.
13. Department of Basic Education. Curriculum and Assessment Policy Statement (CAPS): Life Sciences. Further Education and Training Phase grades 10–12. Pretoria: Government Printing Works; 2011.
14. UMALUSI. What's in the CAPS Package? A comparative study of the National Curriculum Statement (NCS) and the Curriculum and Assessment Policy Statement (CAPS), Further Education and Training (FET) Phase Natural Sciences. Pretoria: Council for Quality Assurance in General and Further Education and Training; 2014.
15. Honey J. Where have all the flowers gone? The place of plants in school science. *J Biol Educ*. 1987;21(3):185–189. <http://dx.doi.org/10.1080/00219266.1987.9654894>
16. Hershey DR. Plant blindness: We have met the enemy and He is Us. *Plant Sci Bull*. 2002;48(3):78–84.

17. National Research Council (US). Plant biology research and training for the 21st century. Committee on Examination of Plant Science Research Programs in the United States. Washington, DC: National Academies Press; 1992. Available from: <http://www.ncbi.nlm.nih.gov/books/NBK234818/>.
18. Bernstein B. The structuring of pedagogic discourse volume IV: Class, codes and control. London: Routledge; 1990. <http://dx.doi.org/10.4324/9780203011263>
19. Taber K. Towards a curricular model of the nature of science. *Sci Educ Netherlands*. 2008;17:179–218. <http://dx.doi.org/10.1007/s11191-006-9056-4>
20. Bernstein B. Pedagogy, symbolic control, and identity: Theory, research, critique. London: Taylor & Francis; 1996. p. 48.
21. Kind V, Taber KS. Science: Teaching school subjects 11–19. London: Routledge; 2005.
22. Schmidt WH, Wang HC, McKnight CC. Curriculum coherence: An examination of the US mathematics and science content standards from an international perspective. *J Curriculum Stud*. 2005;37(5):525–559. <http://dx.doi.org/10.1080/0022027042000294682>
23. American Society of Plant Biologists and ASPB Education Foundation. Principles of plant biology [homepage on the Internet]. c2004 [cited 2015 Feb 20]. Available from: [http://my.aspb.org/?page=EF\\_Principles](http://my.aspb.org/?page=EF_Principles)
24. Khodor J, Halme DG, Walker GC. A hierarchical biology concept framework: A tool for course design. *CBE Life Sci Educ*. 2004;3:111–121. <http://dx.doi.org/10.1187/cbe.03-10-0014>
25. Mauseth JD. Botany: An introduction to plant biology. 5 ed. Burlington, MA: Jones and Bartlett Learning; 2014.
26. Evert R, Eichhorn S. Raven biology of plants. 8th ed. New York: W.H. Freeman & Company; 2012.
27. Stern KR, Jansky S, Bidlack JE. Introductory plant biology. 12th ed. New York: McGraw-Hill Higher Education; 2011.
28. Uno G, Storey R, Moore R. Principles of botany. New York: McGraw-Hill; 2001.
29. Tunnicliffe SD, Ueckert C. Early biology: The critical years for learning. *J Biol Educ*. 2011;45(4):173–175. <http://dx.doi.org/10.1080/00219266.2010.548873>
30. De Beer JJ, Van Wyk B-E. Inheemse kennis in die lewenswetenskapklaskamer: Wetenskap, pseudo-wetenskap, of 'n vermiste skakel? [Indigenous knowledge in the life sciences classroom: Science, pseudo-science or a missing link?]. *S-Afr Tydskr Natuurwet Technol*. 2012;31(1), Art. #368, 5 pages. Afrikaans. <http://dx.doi.org/10.4102/satnt.v31i1.368>
31. Strgar J. Increasing the interest of students in plants. *J Biol Educ*. 2007;42(1):19–23. <http://dx.doi.org/10.1080/00219266.2007.9656102>
32. Schneekloth LH. 'Where did you go?' 'The forest.' 'What did you see?' 'Nothing.' *Children's Environments Q*. 1989;6(1):14–17.
33. Balas B, Momsen JL. Attention 'blinks' differently for plants and animals. *CBE Life Sci Educ*. 2014;13(3):437–443. <http://dx.doi.org/10.1187/cbe.14-05-0080>
34. Nyberg E, Sanders D. Drawing attention to the 'green side of life'. *J Biol Educ*. 2014;48(3):142–153. <http://dx.doi.org/10.1080/00219266.2013.849282>
35. Ebert-May D, Holt E. Seeing the forest and the trees: Research on plant science teaching and learning. *CBE Life Sci Educ*. 2014;13(3):361–362. <http://dx.doi.org/10.1187/cbe.14-06-0105>
36. Hoekstra B. Plant blindness – The ultimate challenge to botanists. *Am Biol Teach*. 2000;62:82–83. [http://dx.doi.org/10.1662/0002-7685\(2000\)062\[0082:PBTU CT\]2.0.CO;2](http://dx.doi.org/10.1662/0002-7685(2000)062[0082:PBTU CT]2.0.CO;2)



# Responses of African elephants towards a bee threat: Its application in mitigating human–elephant conflict

## AUTHORS:

Mduduzi Ndlovu<sup>1,2</sup>  
Emma Devereux<sup>2</sup>  
Melissa Chieffe<sup>2</sup>  
Kendra Asklof<sup>2</sup>  
Alicia Russo<sup>2</sup>

## AFFILIATIONS:

<sup>1</sup>School of Animal, Plant and Environmental Sciences, University of the Witwatersrand, Johannesburg, South Africa

<sup>2</sup>Organisation for Tropical Studies, Skukuza, South Africa

## CORRESPONDENCE TO:

Mduduzi Ndlovu

## EMAIL:

mdu.ndlovu@wits.ac.za

## POSTAL ADDRESS:

School of Animal, Plant and Environmental Sciences, University of the Witwatersrand, Private Bag 3, Wits 2050, South Africa

## DATES:

Received: 10 Feb. 2015

Revised: 22 May 2015

Accepted: 02 June 2015

## KEYWORDS:

*Loxodonta africana*; South Africa; auditory cues; olfactory cues; behaviour

## HOW TO CITE:

Ndlovu M, Devereux E, Chieffe M, Asklof K, Russo A. Responses of African elephants towards a bee threat: Its application in mitigating human–elephant conflict. *S Afr J Sci.* 2016;112(1/2), Art. #2015-0058, 5 pages. <http://dx.doi.org/10.17159/sajs.2016/20150058>

© 2016. The Author(s).  
Published under a Creative Commons Attribution Licence.

Human settlement expansion into elephant ranges, as well as increasing elephant populations within confined areas has led to heightened levels of human–elephant conflict in southern African communities living near protected areas. Several methods to mitigate this conflict have been suggested including the use of bees as an elephant deterrent. We investigated whether bee auditory and olfactory cues (as surrogates for live bees) could be used to effectively deter elephants. We evaluated the responses of elephants in the southern section of the Kruger National Park to five different treatments: (1) control noise, (2) buzzing bee noise, (3) control noise with honey scent, (4) honey scent, and (5) bee noise with honey scent. Elephants did not respond or displayed less heightened responses to the first four treatments. All elephants exposed to the bee noise with honey scent responded with defensive behaviours and 15 out of 21 individuals also fled. We concluded that buzzing bees or honey scent as isolated treatments (as may be the case with dormant beehives) were not effective elephant deterrents, but rather an active beehive emitting a combination of auditory and olfactory cues was a viable deterrent. However, mismatches in the timing of elephant raids and activity of bees may limit the use of bees in mitigating the prevailing human–elephant conflict.

## Introduction

The southern African region accounts for about 40% of the African elephant's (*Loxodonta africana*) total range area.<sup>1–3</sup> Despite measures taken to manage elephant populations in protected areas over the course of the past century, elephant numbers in the region have increased from approximately 170 000 to about 268 000 between 1995 and 2012.<sup>1</sup> As a result of growing African elephant populations within protected areas and increased land cultivation bordering these areas (resulting from human population growth and expansion), there have been numerous reported incidences of human–elephant encounters and conflict, particularly in poor rural farming communities.<sup>4,5</sup> Human–elephant conflict typically refers to interactions between people and elephants that threaten the lives and livelihoods of both parties involved.<sup>6</sup> Beyond growing human and elephant populations, the primary contributing factors to human–elephant conflict in recent years have been increased human settlement and land use change in established elephant migratory corridors. Such human interruptions have in turn affected elephant behaviour and socio-ecology.<sup>6</sup>

Elephant crop raiding is by far the most common cause of human–elephant conflict in South Africa.<sup>4,6</sup> Increased strain on resource availability for growing elephant populations has forced many elephants to leave protected areas and forage on cultivated crops as a means of maximising nutrient intake and reproductive success.<sup>7</sup> Consequently, elephant crop raiding has costs for both humans and elephants. The costs to humans include economic losses through destroyed crops, raided food stores, damaged infrastructure and water sources, and disturbed livestock.<sup>6</sup> Although incidences of damage by elephant crop raiding are low overall, there have been some occurrences of complete crop devastation,<sup>5,6</sup> which can have a substantial influence on the livelihood of the impacted farmer. In some rare cases, crop raiding has also caused injury and loss of human life.<sup>6</sup> Costs to elephants are injury or death at the hands of humans.<sup>6,7</sup>

Considering the high cost of crop raiding for both humans and elephants, a number of deterrent methods have been explored.<sup>6,8,9</sup> Some methods of deterring elephants have included the construction of barriers, translocation and the culling of problem elephants.<sup>8</sup> Although these management strategies have proven effective in some cases, they are often very expensive, beyond the means of most rural communities and can be ethically controversial.<sup>8</sup>

Rural farmers have attempted to defend their crops against elephants using traditional methods, such as lighting fires, making loud noises, and throwing stones.<sup>9</sup> Past research efforts (e.g. Graham and Ochieng<sup>10</sup>) investigated the use of warning alarms, loud noisemakers, watchtowers, spotlights, and African birds eye chillies (*Capsicum frutescens*) in an effort to find an effective deterrent strategy for the management of elephants. Unfortunately, problem elephants would avoid detection by raiding crops at night when people were asleep<sup>7</sup> and hence most deterrent strategies were difficult to implement without constant vigilance. Furthermore, some of these methods have proven to be ineffective and only add costs to farmers, for example, the use of chillies.<sup>11</sup>

Currently, there is a need for an effective, inexpensive, and non-labour intensive method of elephant deterrence for rural communities.<sup>9</sup> Based on evidence of elephants' acute hearing capabilities and sensitive olfactory systems, research has begun to focus on deterrents that target elephants' hearing and smell.<sup>12,13</sup> Additionally, anecdotal evidence shows that despite their thick skin, elephants have sensitive soft regions (i.e. behind the ears, in the eyes, under the trunk and inner-trunk membranes) vulnerable to African honey bee (*Apis mellifera scutellata*) stings.<sup>9,14</sup> There is a report of a bull elephant in Kenya that became permanently blind after being stung by bees multiple times in the eye.<sup>14</sup> King et al.<sup>9</sup> recently compared elephant responses to bee audio recordings and white noise recordings. They found that elephants retreated in response to bee noises and displayed defensive behavioural responses likely to prevent bee stings, including head shaking and dusting, but displayed no significant response to the other noise treatment. This suggests that bee presence could be used as a potential deterrent method for raiding

elephants. However, its long-term effectiveness is still unknown as there is a possibility for animals to get habituated to the threat.

To our knowledge, this type of deterrent has never been explored and evaluated in the high elephant density regions of southern Africa. Therefore, we explored the efficacy of using African honey bee (the bee species found in Kruger National Park) presence to deter elephants. Using African honey bee sound and scent as surrogates for bee presence, we investigated whether bee auditory and olfactory cues could be used to effectively deter elephants. We evaluated the responses of elephants in the southern section of the Kruger National Park in South Africa to five different treatments that were a combination of sound and scent stimuli. Our present study tests the hypothesis that the presence of bees exhibiting a single cue, either olfactory or auditory, is sufficient to deter wild elephants.

## Materials and methods

### Study Site

Data were collected between November 2013 and February 2014 at the height of the summer rainy season in the southern region of the Kruger National Park (hereafter simply referred to as the Kruger) within a 50 km radius from Skukuza rest camp (S 24° 59' 43", E 31°35'34"). The southern section of Kruger is relatively flat and lies in the Lowveld region at altitudes between 200 m and 700 m above mean sea level.<sup>15</sup> The region receives a mean annual rainfall of 500–700 mm,<sup>16</sup> and is characterised by a savanna bushveld dominated by *Acacia spp* trees and high-bulk grasses such as buffalo grass (*Panicum coloratum*), red grass (*Themeda triandra*), and bushveld signal grass (*Urochloa mosabicensis*) species.<sup>16</sup> Large herbivores such as elephant, white rhino (*Ceratotherium simum*), giraffe (*Giraffa camelopardalis*), greater kudu (*Tragelaphus strepsiceros*) and Burchell's zebra (*Equus quagga burchellii*) inhabit the landscape.<sup>17</sup>

Elephant numbers in the Kruger are rapidly increasing. By the end of 2011, there were approximately 14 273 individual elephants in the park.<sup>18</sup> These high densities are reported to negatively alter vegetation structure and diversity in some parts of the park.<sup>19,20</sup>

### Experimental design

We performed preliminary trials in a controlled environment to assess the risk associated with conducting this study in the field. These preliminary experiments involved the exposure of six captive elephants from an elephant sanctuary (S 25°01'39", E 31°07'30") to bee and waterfall audio recordings played from a speaker placed approximately 50 m from the animals. The waterfall and angry buzzing African honey bee noises used in the study were recorded in the Kruger using the high definition voice memo application on an iPhone 4 (Apple Inc, Cupertino, California, USA). Recorded sound treatments were played from an iPhone 4 connected to a 40 watt power Samson Expedition XP40iw rechargeable battery powered wireless PA – Channel 6 (Samson Technologies, Hauppauge, NY, USA) The speaker was placed on top of the research vehicle and sounds were played at maximum volume. Captive elephants ( $n=6$ ) moved away from the bee noises and as expected, did not appear to respond significantly to our control treatment, the waterfall noise. We then shifted our focus to experimentation on wild elephants in the Kruger.

For trials in the Kruger, we drove on management roads around the southern region of the park and arbitrarily selected elephants for observation that were within a 50 m radius of the observation vehicle. Researchers were always accompanied by an armed research assistant. When elephants were located, we observed the animals' pre-stimulus behaviour for 10 min in order to establish a baseline from which we could judge changes in behaviour during the treatment and to ensure that the elephants were somewhat acclimatised to our presence. Most elephants in the Kruger appear to be habituated to the presence of researcher vehicles. We then conducted a behavioural response experiment where we randomly exposed different individuals and groups of wild elephants to one of the five treatments: namely (1) control noise (waterfall), (2) buzzing bee noise, (3) control noise with honey scent, (4) honey scent, and (5) buzzing bee noise with honey scent. The speaker

was positioned on top of the research vehicle, a method similar to that used by McComb.<sup>22</sup> Each treatment was presented for 2 min and the elephant's behavioural response was recorded throughout the duration of the treatment.

We used Estes'<sup>21</sup> behavioural definitions to group and classify 11 possible responses that would indicate the efficacy of a deterrent on a scale of 0–5 (Table 1). Only the highest behavioural response exhibited by each elephant was eventually recorded as that individual's response level for a given trial (Table 1). For each trial, we also recorded total herd size, time of day, ambient temperature, age and sex of the individuals. Age of individual elephants was determined by our experienced game guards and confirmed using Estes.<sup>21</sup> To minimise the chance of subjecting the same elephants to a second treatment on the same day, we only (1) searched each road once a day and (2) selected elephants for observation that were more than 2 km away from the previous herd or individual tested. In instances where we were certain about individual elephant identity, we never sampled those elephants again.

**Table 1:** Classification of elephant behaviour into response levels as adapted from Estes<sup>21</sup>

Response level	Response category	Behaviour
0	No change in pre-stimulus behaviour	No observable reaction
1	Attentive	Listen, freeze, sniff-object
2	Mild disturbance, cautious	Flap ears, reach and touch another elephant
3	Stress, cautious	Temporal gland sweat, grouping
4	Strong deterrence	Panic run, deliberately flee
5	Threat	Charge, mock charge

All sound treatments were kept constant by using the same speaker and volume level. For the honey scent treatment, 50 mL of honey was dissolved in 350 mL of boiling water and the resultant solution was dispersed in a fine mist using a 500 mL handheld plastic spray bottle pointed in the direction of the elephants. We used waterfall noise as our control, based on the assumption that it was a natural and non-threatening sound that would not significantly alter elephant behaviour. Researchers were aware of the treatment being given; however all response levels recorded were dictated by a predefined ethogram (Table 1) and agreed upon by the researchers.

### Data Analysis

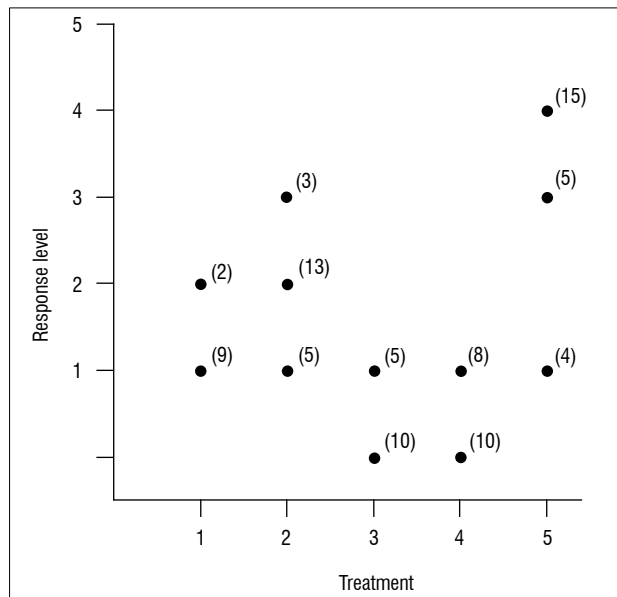
We used the non-parametric Kruskal–Wallis one-way analysis of variance by ranks test with a post-hoc multiple comparison to test for differences in behavioural response levels of elephants to each of the five treatments. A series of Mann–Whitney U tests were used to determine any age-related (adults vs juveniles) differences in behavioural responses for each experimental treatment. All statistical analyses were carried out using the STATISTICA 6<sup>23</sup> program and tested at the 5% level of significance.

## Results

We encountered a total of 136 wild elephants during the study and sampled only 89 individuals, yielding an overall observation rate of 65%. We classed elephants into two age categories: 'Juveniles' ( $n=41$ ) and



'Adults' ( $n=48$ ). We exposed 11 elephants to treatment with control noise alone, 21 elephants to treatment with bee noise alone, 15 elephants to treatment with control noise and honey scent, 18 elephants to treatment with honey scent alone, and 24 elephants to treatment with bee noise and honey scent. The most frequently observed response across all five treatments was an attentive response (response level 1;  $n=27$ ) where subjects would freeze and then raise their trunks to sniff towards the source of the treatment (Figure 1). Level 4 responses, where elephants ran away from the source of the treatment, was the highest behaviour recorded and it was elicited in 15 out of 21 elephants exposed to the bee noise with honey scent treatment (Figure 1). Interestingly, the honey scent alone treatment sometimes (5 out of 15) drew subjects towards the source of the treatment. Three elephants came within 2 m of the vehicle. We classed this behaviour as response level 1.



**Figure 1:** Elephant behavioural response level as a function of experimental treatment. Treatment labels are as follows: (1) control noise, (2) buzzing bee noise, (3) honey scent, (4) control noise with honey scent and (5) bee noise with honey scent. Numbers in parentheses show the sample sizes for each response level.

There was a significant difference in behavioural response levels of elephants to each of the five treatments ( $F=52.15$ ,  $d.f.=4$ ,  $p<0.001$ ,  $n=89$ ). The bee noise and honey scent treatment elicited the highest deterrent behavioural responses from elephants and these response levels were significantly higher than those of the control noise ( $p=0.004$ ), control noise with honey scent ( $p<0.001$ ) and honey scent ( $p<0.001$ ) treatments ( $p<0.001$ ). However, the response levels to the bee noise

with honey scent treatment were not significantly different from the honey scent treatment ( $p=0.126$ ; Table 2). There were no significant differences in the levels of response amongst the (1) control noise, (2) bee noise alone, (3) control noise with honey scent, and (4) honey scent alone treatments (Table 2).

There were no significant differences in behavioural response levels between adults and juveniles within each of the five treatments (control noise:  $U=12.5$ ,  $p=0.788$ ; bee noise;  $U=22$ ,  $p=0.317$ ; control noise with honey scent:  $U=21.5$ ,  $p=0.679$ ; honey scent:  $U=36$ ,  $p=0.762$ ; bee noise with honey scent; control noise with honey scent:  $U=71$ ,  $p=1.000$ ).

Elephants displayed different behavioural responses to each of the various bee threat surrogates and control stimuli presented (Figure 1). Adult and juvenile elephants exhibited a similar within-treatment behavioural response. Mild disturbance behaviour responses and in some instances no observable responses were recorded for all treatment trials except for elephants exposed to the bee noise with honey scent treatment. All elephants exposed to the bee noise with honey scent responded with a cautious behaviour and 15 out of 21 individuals also fled.

## Discussion

The most frequently observed response across all five treatments was an attentive response and it is possible that these elephants were responding to our presence and the vehicle. Cautious responses of elephants to the buzzing bee noise with honey scent as compared with responses to other treatments can be attributed to (1) elephants' equal reliance on both sound and scent as cues for assessing their surroundings<sup>13</sup> and (2) the varying degrees of perceived danger associated with each treatment. Elephants are sensitive to a wide range of sound frequencies aided by their large ear size<sup>12</sup> and they also equally rely on their sense of smell to investigate the environment around them, as the olfactory system in elephants is the primary processing site for chemical stimuli.<sup>13</sup> Another parsimonious explanation, linked to danger perception, is that elephants identified the treatment with the buzzing bee noise only as passing bees, whereas they probably associated the bee noise and honey scent treatment with an active beehive. Distinguishing between a passing swarm and an active beehive seems to have important implications for dictating elephant responses, presumably because of the greater inherent threat posed by encountering a hive as opposed to a passing swarm.<sup>14</sup> Elephants are more at risk of being stung if they come in close contact with a beehive rather than just a passing swarm because African bees are notoriously territorial and have large defensive perimeters surrounding their hives.<sup>14,24</sup> Therefore, elephants might be wary of encountering a large hive where the risk of defensive attack by bees protecting their territory is high.<sup>24</sup>

Contrary to our results that show a minimum response to bee sound alone, King et al.<sup>9</sup> reported that bee sound alone was enough to elicit higher-level responses and also deter elephants in Samburu and Buffalo Springs National Reserves in Kenya. Our results raise new questions

**Table 2:** Post-hoc comparisons of response levels between treatments using the two-tailed Kruskal–Wallis multiple comparisons test. The asterisks (\*) indicate  $p$ -values of treatment responses that were significantly different at a 95% confidence interval

	Treatment			
	Bee noise	Control noise with honey scent	Honey scent	Bee noise with honey scent
Control noise	$p=1.000$	$p=1.000$	$p=0.879$	$p=0.004^*$
Bee noise		$p=0.052$	$p=0.002^*$	$p=0.126$
Control noise with honey scent			$p=1.000$	$p<0.001^*$
honey scent				$p<0.001^*$

about what makes elephants in the Kruger different from elephants in other places. One possible explanation for our results is that the high density of elephants in the Kruger<sup>18</sup> compared to elephant densities in Kenyan reserves<sup>1</sup> increases the likelihood of bee encounters in the South African park. As a consequence of the increased probability of elephant exposure to bees in the Kruger, the subjects in our study were potentially familiarised with the sound and smell of live beehives, and therefore better equipped with cues that indicate a realistic bee threat.

We concluded that the observed cautious behavioural responses from elephants when exposed to our bee threat proxies provide strong support for our hypothesis that these same elephants would be deterred by a live bee threat. It therefore suggests that a treatment evoking both the olfactory and auditory cues of a 'bee threat' is required to deter wild elephants. The effectiveness of the mixed stimulus treatment can be explained by the fact that the bee noise and honey scent treatment better imitated the presence of an active beehive than the treatment with the bee noise alone. The combination of sound and scent was a more realistic representation of a bee threat, which elicited a greater response from the elephants because: (1) elephants likely associated this combined stimulus with the presence of an active beehive, which indicated a greater threat to elephants than the sound of a passing swarm; and (2) elephants rely on both auditory and olfactory cues to detect a threat.

Studies in Kenya have demonstrated the effectiveness of using beehive fences to deter elephants from raiding farms and damaging large trees.<sup>9,14</sup> Our findings indicate that a similar innovation could also be used to mitigate the human–elephant conflict on farms and in settlements surrounding the Kruger (both in South Africa and Mozambique) and other parts of southern Africa. In addition to aiding in human–elephant conflict mitigation, apiculture (beekeeping) has potential benefits for sustainable community-based conservation, particularly because honey harvesting is a traditional practice in many African cultures.<sup>25</sup> Aside from potentially reducing losses from elephant raids, apiculture can provide employment and income opportunities for communities through the production of marketable products such as honey and wax.<sup>25,26</sup>

However, our study also points out one key limitation to the use of bees to deter elephants. We know that elephant raids in most parts of southern Africa (1) occur at night when temperatures are low and (2) are prominent in winter when natural browse and graze opportunities are at their minima.<sup>6</sup> Unfortunately, most African honey bees tend to be dormant (less active) at night and when temperatures are low. Our findings therefore imply that the use of active (buzzing and scent emitting) bees as recommended by King et al.<sup>9</sup>, may be seriously mismatched with the timing of elephant raids. Perhaps the development of some trigger mechanism to activate dormant bees when elephant raids occur will remedy the problem.

## Acknowledgements

We thank the Organisation for Tropical Studies and South African National Parks (SANParks) for supporting this research. We are grateful to Ceinwin Smith, Karen Vickers, Dax Mackay, Philip Mhlava, Donovan Tye, Laurence Kruger, Alyssa Browning, Tyler Maddox, Cassandra Pestana and David Purdy for their assistance with data collection. Lastly, we wish to thank André Kotzé, the rest of the staff at Elephant Whispers Sanctuary and the SANParks research assistants. Their time and helpful insight into elephant behaviour is much appreciated.

## Authors' Contributions

M.N. designed the research. E.D., M.C., K.A., and A.R. collected and analysed the data. All authors contributed equally to the writing of the manuscript.

## References

1. International Union for the Conservation of Nature Species Survival Commission: African Elephant Specialist Group. Elephant Database: Regional totals for southern Africa for 2013 [document on the Internet]. c2013 [cited 2014 November 28]. Available from: [http://www.elephantdatabase.org/preview\\_report/2013\\_africa/Loxodonta\\_africana/2012/Africa/Southern\\_Africa](http://www.elephantdatabase.org/preview_report/2013_africa/Loxodonta_africana/2012/Africa/Southern_Africa)


2. Blanc J. *Loxodonta africana*. International Union for the Conservation of Nature (IUCN): IUCN red list of threatened species. Version 2013.1; 2008 [document on the Internet]. c2014 [cited 2014 November 25]. Available from: <http://www.iucnredlist.org>. <http://dx.doi.org/10.2305/IUCN.UK.2008.RLTS.T12392A3339343.en>
3. Carruthers J, Boshoff A, Slotow R, Biggs HC, Avery G, Matthews W. The elephant in South Africa: History and distribution. In: Scholes RJ, Mennel KG, editors. The 2007 scientific assessment of elephant management in South Africa. Johannesburg: Wits University Press; 2008. p. 1–40.
4. Lee PC, Graham MD. African elephants (*Loxodonta africana*) and human–elephant interactions: Implications for conservation. *Int Zoo Yearb*. 2006;40:9–19. <http://dx.doi.org/10.1111/j.1748-1090.2006.00009.x>
5. Graham MD, Notter B, Adams WM, Lee PC, Ochieng TN. Patterns of crop raiding by elephants, *Loxodonta africana*, in Laikipia, Kenya, and the management of human–elephant conflict. *Syst Biodivers*. 2009;8:435–445. <http://dx.doi.org/10.1080/14772000.2010.533716>
6. Twine W, Magome H. Interactions between elephant and people. In: Scholes RJ, Mennel KG, editors. The 2007 scientific assessment of elephant management in South Africa. Johannesburg: Wits University Press; 2008. p. 148–168.
7. Chiyo PI, Moss CJ, Alberts CS. The influence of life history milestones and association networks on crop-raiding behavior in male African elephants. *PLoS ONE*. 2012;7:1–11. <http://dx.doi.org/10.1371/journal.pone.0031382>
8. Graham MD, Nyumna TO, Kahiro G, Ngotho M, Adams WM. Trials of farm based deterrents to mitigate crop-raiding by elephants adjacent to the Rumuruti Forest in Laikipia, Kenya. Nanyuki: Laikipia Elephant Project; 2009.
9. King, LE, Soltis J, Douglas-Hamilton I, Savage A, Vollrath F. Bee threat elicits alarm call in African elephants. *PLoS ONE*. 2010;5:1–9. <http://dx.doi.org/10.1371/journal.pone.0010346>
10. Graham MD, Ochieng, T. Uptake and performance of farm-based measures for reducing crop-raiding by elephants *Loxodonta africana* among smallholder farms in Laikipia District, Kenya. *Oryx*. 2008;42:76–82. <http://dx.doi.org/10.1017/S0030605308000677>
11. Hedges S, Gunaryadi D. Reducing human–elephant conflict: Do chillies help deter elephants from entering crop fields. *Oryx*. 2009;44:139–146. <http://dx.doi.org/10.1017/S0030605309990093>
12. Reuter T, Nummela S. Elephant hearing. *J Acoust Soc Am*. 1998;104:1122–1123 <http://dx.doi.org/10.1121/1.423341>
13. Ngwenya A, Patzke N, Ilhunwo AO, Manger PR. Organisation and chemical neuroanatomy of the African elephant (*Loxodonta africana*) olfactory bulb. *Brain Struct Funct*. 2011;216:403–416. <http://dx.doi.org/10.1007/s00429-011-0316-y>
14. Vollrath F, Douglas-Hamilton I. African bees to control African elephants. *Naturwissenschaften*. 2002;89:508–511. <http://dx.doi.org/10.1007/s00114-002-0375-2>
15. Mabunda D, Pienaar DJ, Verhoef J. In: Du Toit JT, Rogers KH, Biggs HC, editors. The Kruger experience: Ecology and management of savanna heterogeneity. Washington, DC: Island Press; 2003. p. 3–21.
16. Venter FJ, Scholes RJ, Eckhardt HC. The abiotic template and its associated vegetation pattern. In: Du Toit JT, Rogers KH, Biggs HC, editors. The Kruger experience: Ecology and management of savanna heterogeneity. Washington, DC: Island Press; 2003. p. 83–129.
17. Owen-Smith N, Ogutu J. Rainfall influences on ungulate population dynamics. In: Du Toit JT, Rogers KH, Biggs HC, editors. The Kruger experience: Ecology and management of savanna heterogeneity. Washington, DC: Island Press; 2003. p. 310–331.
18. South African National Parks (SANParks). Estimates for animal abundances in parks in the northern region. South African National Parks annual report 2011/2012. Pretoria: SanParks. Available from: <http://www.sanparks.co.za/assets/docs/general/annual-report-2012.pdf>; 2012
19. Grainger M, Van Aarde RJ, Whyte I. Landscape heterogeneity and the use of space by elephants in the Kruger National Park, South Africa. *Afr J Ecol*. 2005;43:369–375. <http://dx.doi.org/10.1111/j.1365-2028.2005.00592.x>

20. Whyte I, Van Aarde R, Pimm SL. Managing the elephants of Kruger National Park. *Anim Conserv*. 1998;1:77–83. <http://dx.doi.org/10.1111/j.1469-1795.1998.tb00014.x>
21. Estes RD. Elephants: Order Proboscidea, Family Elephantidae. In: Estes RD, editor. *The Behavior Guide to African Mammals*. Johannesburg: Russel Friedman Books; 1991. p. 259–267.
22. McComb K, Shannon G, Durant SM, Sayialel S, Slotow R, Poole J, Moss C. Leadership in elephants: The adaptive value of age. *Proc R Soc B*. 2011;278:3270–3276. <http://dx.doi.org/10.1098/rspb.2011.0168>
23. StatSoft, Inc. STATISTICA. Version 6.0. Tulsa, OK, USA; 2006.
24. Breed MD, Guzman-Novoa E, Hunt GJ. Defensive behavior of honey bees: Organization, genetics, and comparisons with other bees. *Ann Rev Entomol*. 2004;49:217–298. <http://dx.doi.org/10.1146/annurev.ento.49.061802.123155>
25. King LE. The interaction between the African elephant (*Loxodonta africana*) and the African honey bee (*Apis mellifera scutellata*) and its potential applications as an elephant deterrent [MSc dissertation]. Oxford: University of Oxford; 2010.
26. Hill DB, Webster TC. Apiculture and forestry (bees and trees). *Agroforest Syst*. 1995;29:313–320. <http://dx.doi.org/10.1007/BF00704877>



# High-speed detection of emergent market clustering via an unsupervised parallel genetic algorithm

## AUTHORS:

Dieter Hendricks<sup>1</sup> 

Tim Gebbie<sup>1</sup> 

Diane Wilcox<sup>1</sup> 

## AFFILIATION:

<sup>1</sup>School of Computer Science and Applied Mathematics, University of the Witwatersrand, Johannesburg, South Africa

## CORRESPONDENCE TO:

Dieter Hendricks

## EMAIL:

dieter.hendricks@students.wits.ac.za

## POSTAL ADDRESS:

School of Computer Science and Applied Mathematics, University of the Witwatersrand, Private Bag, Wits 2050, South Africa

## DATES:

**Received:** 01 Oct. 2014

**Revised:** 12 May 2015

**Accepted:** 24 June 2015

## KEYWORDS:

unsupervised clustering; genetic algorithms; parallel algorithms; financial data processing; maximum likelihood clustering

## HOW TO CITE:

Hendricks D, Gebbie T, Wilcox D. High-speed detection of emergent market clustering via an unsupervised parallel genetic algorithm. *S Afr J Sci.* 2016;112(1/2), Art. #2014-0340, 9 pages. <http://dx.doi.org/10.17159/sajs.2016/20140340>

We implement a master-slave parallel genetic algorithm with a bespoke log-likelihood fitness function to identify emergent clusters within price evolutions. We use graphics processing units (GPUs) to implement a parallel genetic algorithm and visualise the results using disjoint minimal spanning trees. We demonstrate that our GPU parallel genetic algorithm, implemented on a commercially available general purpose GPU, is able to recover stock clusters in sub-second speed, based on a subset of stocks in the South African market. This approach represents a pragmatic choice for low-cost, scalable parallel computing and is significantly faster than a prototype serial implementation in an optimised C-based fourth-generation programming language, although the results are not directly comparable because of compiler differences. Combined with fast online intraday correlation matrix estimation from high frequency data for cluster identification, the proposed implementation offers cost-effective, near-real-time risk assessment for financial practitioners.

## Introduction

Advances in technology underpinning multiple domains have increased the capacity to generate and store data and metadata relating to domain processes. The field of data science is continuously evolving to meet the challenge of gleaning insights from these large data sets, with extensive research in exact algorithms, heuristics and meta-heuristics for solving combinatorial optimisation problems. The primary advantage of using exact methods is the guarantee of finding the global optimum for the problem. However, a disadvantage when solving complex (NP-hard) problems is the exponential growth of the execution time proportional to the problem instance size.<sup>1</sup> Heuristics tend to be efficient, but solution quality cannot be guaranteed and techniques often are not versatile.<sup>2</sup> Meta-heuristics attempt to consolidate these two approaches and deliver an acceptable solution in a reasonable time frame. A large number of meta-heuristics designed for solving complex problems exist in the literature and the genetic algorithm (GA) has emerged as a prominent technique, using intensive global search heuristics that explore a search space intelligently to solve optimisation problems.

Although the algorithms must traverse large spaces, the computationally intensive calculations can be performed independently. Compute Unified Device Architecture (CUDA) is Nvidia's parallel computing platform which is well suited to many computational tasks, particularly those for which data parallelism is possible. Implementing a GA to perform cluster analysis on vast data sets using this platform allows one to mine through the data relatively quickly and at a fraction of the cost of those of large data centres or computational grids.

A number of authors have considered parallel architectures to accelerate GAs.<sup>3-10</sup> While the work of Kromer et al.<sup>7</sup> is conceptually similar to the implementation proposed in this paper, a key difference is our choice of fitness function for the clustering scheme.

Giada and Marsili<sup>11</sup> propose an unsupervised, parameter-free approach to finding data clusters, based on the maximum likelihood principle. They derive a log-likelihood function, where a given cluster configuration can be assessed to determine whether it represents the inherent structure for the data set: cluster configurations which approach the maximum log-likelihood are better representatives of the data structure. This log-likelihood function is thus a natural candidate for the fitness function in a GA implementation, where the population continually evolves to produce a cluster configuration which maximises the log-likelihood. The optimal number of clusters is a free parameter, unlike in traditional techniques where the number of clusters needs to be specified *a priori*. While unsupervised approaches have been considered (see Omran et al.<sup>12</sup> and references therein), the advantage of the Giada and Marsili approach is that it has a natural interpretation for clustering in the application domain explored here.

Monitoring intraday clustering of financial instruments allows one to better understand market characteristics and systemic risks. While GAs provide a versatile methodology for identifying such clusters, serial implementations are computationally intensive and can take a long time to converge to a best approximation. In this paper, we introduce a maintainable and scalable master-slave parallel genetic algorithm (PGA) framework for unsupervised cluster analysis on the CUDA platform, which is able to detect clusters using the Giada and Marsili likelihood function. Applying the proposed cluster analysis approach and examining the clustering behaviour of financial instruments, offers a unique perspective to monitor the intraday characteristics of the stock market and the detection of structural changes in near real time. The novel implementation presented in this paper builds on the contribution of Cieslakiewicz<sup>13</sup>. While we provide an overview and specific use-case for the algorithm in this paper, we also are investigating aspects of adjoint parameter tuning, performance scalability and the impact on solution quality for varying stock universe sizes and cluster types.

## Cluster analysis

Cluster analysis groups objects according to metadata describing the objects or their associations.<sup>14</sup> The goal is to ensure that objects within a group exhibit similar characteristics and are unrelated to objects in other groups. The greater the homogeneity within a group, and the greater the heterogeneity between groups, the more



pronounced the clustering. In order to isolate clusters of similar objects, one needs to utilise a data clustering approach that will recover inherent structures efficiently.

### The correlation measure of similarity

The correlation measure is an approach to standardise data by using the statistical interdependence between data points. Correlation indicates the direction (positive or negative) and the degree or strength of the relationship between two data points. The most common correlation coefficient which measures the relationship between data points is the Pearson correlation coefficient, which is sensitive only to a linear relationship between points. The Pearson correlation is +1 in the case of a perfect positive linear relationship, -1 in the case of a perfect negative linear relationship and some value between -1 and +1 in all other cases, with values close to 0 signalling negligible interdependence.

### Clustering procedures

Any specific clustering procedure entails optimising some kind of criterion, such as minimising the within-cluster variance or maximising the distance between the objects or clusters.

#### Cluster analysis based on the maximum likelihood principle

Maximum likelihood estimation is a method of estimating the parameters of a statistical model. Data clustering on the other hand deals with the problem of classifying or categorising a set of  $N$  objects or clusters, so that the objects within a group or cluster are more similar than objects belonging to different groups. If each object is identified by  $D$  measurements, then an object can be represented as a tuple,  $\bar{x}_i = (x_i^{(1)}, \dots, x_i^{(D)})$ ,  $i=1, \dots, N$  in a  $D$ -dimensional space. Data clustering is used to identify clusters as more densely populated regions in this vector space. Thus, a configuration of clusters is represented by a set  $S = \{s_1, \dots, s_M\}$  of integer labels, where  $s_i$  denotes the cluster to which object  $i$  belongs and  $N$  is the number of objects<sup>11</sup> (if  $s_i = s_j = s$ , then object  $i$  and object  $j$  reside in the same cluster), and if  $s_i$  takes on values from 1 to  $M$  and  $M=N$ , then each cluster is a singleton cluster constituting one object only.

#### Analogy to the Potts model

One can apply super-paramagnetic ordering of a  $q$ -state Potts model directly for cluster identification.<sup>15</sup> In a market Potts model, each stock can take on  $q$ -states and each state can be represented by a cluster of similar stocks.<sup>15-17</sup> Cluster membership is indicative of some commonality among the cluster members. Each stock has a component of its dynamics as a function of the state it is in and a component of its dynamics influenced by stock specific noise. In addition, there may be global couplings that influence all the stocks, i.e. the external field that represents a market mode.

In the super-paramagnetic clustering approach, the cost function can be considered as a Hamiltonian function whose low energy states correspond to cluster configurations that are most compatible with the data sample. Structures are then identified with configurations  $S = \{s_i\}_{i=1}^N$  for the cluster indices  $s_i$  which represent the cluster to which the  $i$ th object belongs. This allows one to interpret  $s_i$  as a Potts spin in the Potts model Hamiltonian with  $J_{ij}$  decreasing with the distance between objects.<sup>15,16</sup> The Hamiltonian takes on the form:

$$H_g = - \sum_{s_i, s_j \in S} J_{ij} \delta(s_i, s_j) - \frac{1}{\beta} \sum_i h_i^m s_i, \quad \text{Equation 1}$$

where the spins  $s_j$  can take on  $q$ -states and the external magnetic fields are given by  $h_i^m$ . The first term represents common internal influences and the second term represents external influences. We ignore the second term when fitting data, as we include shared factors directly in later sections when we discuss information and risk and the influence of these on price changes.

In the Potts model approach, one can think of the coupling parameters  $J_{ij}$  as being a function of the correlation coefficient.<sup>16,17</sup> The coupling parameters are used to specify a distance function that decreases with distance between objects. If all the spins are related in this way then each

pair of spins is connected by some non-vanishing coupling,  $J_{ij} = J_{ij}(c_{ij})$ . In this model, the case where there is only one cluster can be thought of as a ground state. As the system becomes more excited, it could break up into additional clusters and each cluster would have specific Potts magnetisations, even though net magnetisation may remain zero for the complete system. Generically, the correlation would then be both a function of time and temperature in order to encode both the evolution of clusters as well as the hierarchy of clusters as a function of temperature. In the basic approach, one is looking for the lowest energy state that fits the data. In order to parameterise the model efficiently one can choose to make the Noh<sup>18</sup> ansatz and use this to develop a maximum-likelihood approach of Giada and Marsili<sup>17</sup> rather than explicitly solving the Potts Hamiltonian numerically<sup>15,16</sup>.

#### Giada and Marsili clustering technique

Following Giada and Marsili<sup>17</sup>, we assume that price increments evolve under Noh<sup>18</sup> model dynamics, whereby objects belonging to the same cluster should share a common component:

$$\bar{x}_i = g_{s_i} \bar{\eta}_{s_i} + \sqrt{1-g_{s_i}^2} \bar{\epsilon}_i. \quad \text{Equation 2}$$

Here,  $\bar{x}_i$  represents the features of object  $i$  and  $s_i$  is the label of the cluster to which the object belongs. The data have been normalised to have zero mean and unit variance.  $\bar{\epsilon}_i$  is a vector describing the deviation of object  $i$  from the features of cluster  $s$  and includes measurement errors, while  $\bar{\eta}_{s_i}$  describes cluster-specific features.  $g_s$  is a loading factor that emphasises the similarity or difference between objects in cluster  $s$ . In this research the data set refers to a set of the objects, denoting  $N$  assets or stocks, and their features are prices across  $D$  days in the data set. The variable  $i$  is indexing stocks or assets, whilst  $d$  is indexing days.

If  $g_s = 1$ , all objects with  $s_i = s$  are identical, whilst if  $g_s = 0$ , all objects are different. The range of the cluster index is from 1 to  $N$  in order to allow for singleton clusters of one object or asset each.

If one takes Equation 2 as a statistical hypothesis and assumes that both  $\bar{\eta}_{s_i}$  and  $\bar{\epsilon}_i$  are Gaussian vectors with zero mean and unit variance, for values of  $i, s = 1, \dots, N$ , it is possible to compute the probability density  $P(\{\bar{x}_i\} | G, S)$  for any given set of parameters  $(G, S) = (\{g_s\}, \{s_i\})$  by observing the data set  $\{x_i\}, i, s = 1, \dots, N$  as a realisation of the common component of Equation 2 as follows<sup>11</sup>:

$$P(\{\bar{x}_i\} | G, S) = \prod_{d=1}^D \left\langle \prod_{i=1}^N \delta(x_i(d) - g_{s_i} \bar{\eta}_{s_i} + \sqrt{1-g_{s_i}^2} \bar{\epsilon}_i) \right\rangle. \quad \text{Equation 3}$$

The variable  $\delta$  is the Dirac delta function and  $\langle \dots \rangle$  denotes the mathematical expectation. For a given cluster structure  $S$ , the likelihood is maximal when the parameter  $g_s$  takes the values

$$g_s^* = \begin{cases} \sqrt{\frac{c_s - n_s}{n_s^2 - n_s}} & \text{for } n_s > 1, \\ 0 & \text{for } n_s \leq 1. \end{cases} \quad \text{Equation 4}$$

The quantity  $n_s$  in Equation 4 denotes the number of objects in cluster  $s$ , i.e.

$$n_s = \sum_{i=1}^N \delta_{s_i s}. \quad \text{Equation 5}$$

The variable  $c_s$  is the internal correlation of the  $s$ th cluster, denoted by:

$$c_s = \sum_{i=1}^N \sum_{j=1}^N C_{i,j} \delta_{s_i s} \delta_{s_j s}. \quad \text{Equation 6}$$

The variable  $C_{i,j}$  is the Pearson correlation coefficient of the data, denoted by:

$$C_{i,j} = \frac{\bar{x}_i \bar{x}_j}{\sqrt{|\bar{x}_i|^2} \sqrt{|\bar{x}_j|^2}}. \quad \text{Equation 7}$$

The maximum likelihood of structure  $S$  can be written as  $P(G^*, S | \bar{X}) \propto \exp^{L_c(S)}$  (see Sornette<sup>19</sup>), where the resulting likelihood function per feature  $L_c$  is denoted by:

$$L_c(S) = \frac{1}{2} \sum_{s: n_s > 1} \left( \log \frac{n_s}{c_s} + (n_s - 1) \log \frac{n_s^2 - n_s}{n_s^2 - c_s} \right). \quad \text{Equation 8}$$

From Equation 8, it follows that  $L_c=0$  for clusters of objects that are uncorrelated, i.e. where  $g_s^* = 0$  or  $c_s = n_s$  or when the objects are grouped in singleton clusters for all the cluster indexes ( $n_s=1$ ). Equation 8 illustrates that the resulting maximum likelihood function for  $S$  depends on the Pearson correlation coefficient  $C_{ij}$  and hence exhibits the following advantages in comparison to conventional clustering methods:

- It is *unsupervised*: The optimal number of clusters is unknown *a priori* and not fixed at the beginning
- The interpretation of results is *transparent* in terms of the model, namely Equation 2.

Giada and Marsili<sup>11</sup> propose that  $\max_s L_c(S)$  provides a measure of structure inherent in the cluster configuration represented by the set  $S = \{s_1, \dots, s_n\}$ . The higher the value, the more pronounced the structure.

## Parallel genetic algorithms

In order to localise clusters of normalised stock returns in financial data, Giada and Marsili made use of a simulated annealing algorithm,<sup>11,17</sup> with  $-L_c$  as the cost function for their application on real-world data sets to substantiate their approach. This simulated annealing algorithm was then compared to other clustering algorithms, such as K-means, single linkage, centroid linkage, average linkage, merging and deterministic maximisation.<sup>11</sup> The technique was successfully applied to South African financial data by Mbambiso<sup>20</sup>, using a serial implementation of a simulated annealing algorithm<sup>20,21</sup>.

Simulated annealing and deterministic maximisation provided acceptable approximations to the maximum likelihood structure, but were inherently computationally expensive. We promote the use of PGAs as a viable approach to approximate the maximum likelihood structure.  $L_c$  will be used as the fitness function and a PGA algorithm will be used to find the maximum for  $L_c$ , in order to efficiently isolate clusters in correlated financial data.

### GA principle and genetic operators

One of the key advantages of GAs is that they are conceptually simple. The core algorithm can be summarised into the following steps: (1) initialise population, (2) evolve individuals, (3) evaluate fitness and (4) select individuals to survive to the next generation. GAs exhibit the trait of broad applicability,<sup>22</sup> as they can be applied to any problem whose solution domain can be quantified by a function which needs to be optimised.

Specific genetic operators are applied to the parents, in the process of reproduction, which then give rise to offspring. The genetic operators can be classified as follows:

**Selection:** The purpose of selection is to isolate fitter individuals in the population and allow them to propagate in order to give rise to new offspring with higher fitness values. We implemented the *stochastic universal sampling selection operator*, in which individuals are mapped to contiguous segments on a line in proportion to their fitness values.<sup>23</sup> Individuals are then selected by sampling the line at uniformly spaced intervals. Although fitter individuals have a higher probability of selection, this technique improves the chances that weaker individuals will be selected, allowing diversity to enter the population and reducing the probability of convergence to a local optimum.

**Crossover:** Crossover is the process of mating two individuals, with the expectation that they can produce a fitter offspring.<sup>22</sup> The crossover genetic operation involves the selection of random loci to mark a cross site within the two parent chromosomes, copying the genes to

the offspring. A bespoke *knowledge-based crossover operator*<sup>13</sup> was developed for our implementation, in order to incorporate domain knowledge and improve the rate of convergence.

**Mutation:** Mutation is the key driver of diversity in the candidate solution set or search space.<sup>22</sup> It is usually applied after crossover and aims to ensure that genetic information is randomly distributed, in order to avoid convergence to local minima. It introduces new genetic structures in the population by randomly modifying some of its building blocks and enables the algorithm to traverse the search space globally.

**Elitism:** Coley<sup>24</sup> states that fitness-proportional selection does not necessarily favour the selection of any particular individual, even if it is the fittest. Thus, the fittest individuals may not survive an evolutionary cycle. Elitism is the process of preserving the fittest individuals by direct promotion to the next generation, without any genetic transformations due to crossover or mutation.<sup>22</sup>

**Replacement:** Replacement is the last stage of any evolution cycle, in which the algorithm replaces old members of the current population with new members.<sup>22</sup> This mechanism ensures that the population size remains constant, while the weakest individuals in each generation are dropped.

Although GAs are very effective for solving complex problems, this positive trait can unfortunately be offset by long execution times caused by the traversal of the search space. GAs lend themselves to parallelisation, provided the fitness values can be determined independently for each of the candidate solutions. While a number of schemes have been proposed in the literature to achieve this parallelisation,<sup>8,22,25</sup> we have chosen to implement the *master-slave* model.

### Master-slave parallelisation

Master-slave PGAs, also denoted as global PGAs, involve a single population, distributed amongst multiple processing units for determination of fitness values and the consequent application of genetic operators. They allow for computation on shared-memory processing entities or any type of distributed system topology, for example grid computing.<sup>8</sup>

Ismail<sup>25</sup> provides a summary of the key features of the master-slave PGA: the algorithm uses a single population (stored by the master) and the fitness evaluation of all of the individuals is performed in parallel (by the slaves). Communication occurs only as each slave receives an individual (or subset of individuals) to evaluate and when the slaves return the fitness values, sometimes after mutation has been applied with the given probability. The particular algorithm implemented in this paper is synchronous, i.e. the master waits until it has received the fitness values for all individuals in the population before proceeding with selection and mutation. The synchronous master-slave PGA thus has the same properties as a conventional GA, except evaluation of the fitness of the population is achieved at a faster rate. The algorithm is relatively easy to implement and a significant speed-up can be expected if the communications cost does not dominate the computation cost. The whole process has to wait for the slowest processor to finish its fitness evaluations until the selection operator can be applied.

A number of authors have used the Message Parsing Interface (MPI) paradigm to implement a master-slave PGA. Digalakis and Margaritis<sup>26</sup> implement a synchronous MPI PGA and shared-memory PGA, whereby fitness computations are parallelised and other genetic operators are applied by the master node only. They demonstrate a computation speed-up which scales linearly with the number of processors for large population sizes. Zhang et al.<sup>27</sup> use a centralised control island model to concurrently apply genetic operators to sub-groups, with a bespoke migration strategy using elite individuals from sub-groups. Nan et al.<sup>28</sup> used the MATLAB parallel computing and distributed computing toolboxes to develop a master-slave PGA, demonstrating its efficacy on the image registration problem when using a cluster computing configuration.

For our implementation, we made use of the Nvidia CUDA platform to achieve massive parallelism by utilising the graphical processing unit (GPU) streaming multiprocessors (SM) as slaves, and the central processing unit (CPU) as master.

## Computational platform and implementation

CUDA is Nvidia’s platform for massively parallel high-performance computing on the Nvidia GPUs. At its core are three key abstractions: a hierarchy of thread groups, shared memories and barrier synchronisation. Full details on the execution environment, thread hierarchy, memory hierarchy and thread synchronisation schemes have been omitted here, but we refer the reader to Nvidia technical documentation<sup>29,30</sup> for a comprehensive discussion.

### Specific computational environment

The CUDA algorithm and the respective testing tools were developed using Microsoft Visual Studio 2012 Professional, with the Nvidia Nsight extension for CUDA-C projects. The configurations shown in Table 1 were tested to determine the versatility of the CUDA clustering algorithms.

We had the opportunity to test two candidate graphics cards for the algorithm implementation: the Nvidia GTX Titan Black and the Nvidia TESLA C2050. Both cards offer double-precision calculations and a similar number of CUDA cores and TFLOPS (tera floating point operations per second); however, the GTX card is significantly cheaper than the TESLA card. The primary reason for this cost difference is the use of ECC (error check and correction) memory on the TESLA cards, where extra memory bits are present to detect and fix memory errors.<sup>31</sup> The presence of ECC memory ensures consistency in results generated from the TESLA card, which is critical for rigorous scientific computing. In further investigations, we will explore the consistency of the solution quality generated from the GTX card, and whether the resultant error is small enough to justify the cost saving compared to the TESLA card.

### Implementation

The following objectives were considered in this study: (1) to investigate and tune the behaviour of the PGA implementation using a pre-defined set of 40 simulated stocks, featuring four distinct and disjoint clusters, (2) to identify clusters in a real-world data set, namely high-frequency price evolutions of stocks and (3) to test the efficiency of the GPU environment.

### Representation

We used integer-based encoding for the representation of individuals in the GA:

$$\text{Individual} = S = \{s_1, s_2, \dots, s_{i-1}, s_i, \dots, s_N\} \quad \text{Equation 9}$$

where  $s_i = 1, \dots, K$  and  $i = 1, \dots, N$ . Here  $s_i$  is the cluster to which object  $i$  belongs. In terms of the terminology pertaining to GAs, the  $i$ th gene denotes the cluster to which the  $i$ th object or asset belongs. The numbers of objects or assets is  $N$ ; thus to permit the possibility of an all-singleton configuration, we let  $K=N$ . This representation was implemented by Gebbie et al.<sup>21</sup> in their serial GA and was adopted in this research.

### Fitness function

The Giada and Marsili maximum log-likelihood function  $L_c$ , as shown in Equation 8, was used as the fitness function. This function is used to determine whether the cluster configuration represents the inherent structure of the data set, i.e. it is used to detect if the GA converges to the fittest individual, corresponding to the cluster configuration which best explains the structure amongst correlated assets or objects in the data set.

### Master-slave implementation

The unparallelised MATLAB GA implementation of the clustering technique by Gebbie et al.<sup>21</sup> served as a starting point. In order to maximise the performance of the GA, the application of genetic operators and evaluation of the fitness function were parallelised for the CUDA framework.<sup>13</sup> A summarised exposition is presented here.

Emphasis was placed on outsourcing as much of the GA execution to the GPU and using GPU memory as extensively as possible.<sup>32</sup> The master-slave PGA uses a single population, for which evaluation of the individuals and successive application of genetic operators are conducted in parallel. The global parallelisation model does not predicate anything about the underlying computer architecture, so it can be implemented efficiently on a shared-memory and distributed-memory model platform.<sup>22</sup> Delegating these tasks to the GPU and making extensive use of GPU memory minimises the data transfers between the host and device. These transfers have a significantly lower bandwidth than data transfers between shared or global memory and the kernel executing on the GPU. The algorithm in Gebbie et al.<sup>21</sup> was modified to maximise the performance of the master-slave PGA and to have a clear distinction between the master node (CPU) and slave nodes (GPU streaming multiprocessors). The CPU controls the evolutionary process by issuing the commands for the GA operations to be performed by

Table 1: Development, testing and benchmarking environments

Environment	Configuration	Framework
WINDOWS_GTX_CUDA	Windows 7 Professional Service Pack 1 (64-bit) Core i7-4770K CPU@3.50 GHz x 8, 32 GB RAM Nvidia GTX Titan Black with 6 GB RAM, CC: 3.0, SM: 3.5	CUDA-C 5.5 (parallel)
WINDOWS_GTX_MATLAB	Windows 7 Professional Service Pack 1 (64-bit) Core i7-4770K CPU@3.50 GHz x 8, 32 GB RAM Nvidia GTX Titan Black with 6 GB RAM, CC: 3.0, SM: 3.5	MATLAB 2013a (serial)
WINDOWS_TESLA_CUDA	Windows 7 Professional Service Pack 1 (64-bit) Intel Core i7-X980 CPU@3.33 GHz x 12, 24 GB RAM Nvidia TESLA C2050 with 2.5 GB RAM, CC: 2.0, SM: 2.0	CUDA-C 5.5 (parallel)
WINDOWS_TESLA_MATLAB	Windows 7 Professional Service Pack 1 (64-bit) Intel Core i7-X980 CPU@3.33 GHz x 12, 24 GB RAM Nvidia TESLA C2050 with 2.5 GB RAM, CC: 2.0, SM: 2.0	MATLAB 2013a (serial)

CPU, central processing unit; RAM, random access memory; CC, compute capability; SM, streaming multiprocessor.

the GPU streaming multiprocessors. The pseudo-code for the algorithm implemented is shown in Algorithm 1.

**Algorithm 1:** Master-slave parallel genetic algorithm for cluster identification

```

Initialise ecosystem for evolution
Size the thread blocks and grid to achieve greatest parallelisation
ON GPU: Create initial population
while TRUE do
ON GPU: Evaluate fitness of all individuals
ON GPU: Evaluate state and statistics
ON GPU: Determine if termination criteria are met
if YES then
  Terminate ALGO; exit while loop;
else
  Continue
end if
ON GPU: Isolate fittest individuals
ON GPU: Apply elitism
ON GPU: Apply scaling
ON GPU: Apply genetic operator: selection
ON GPU: Apply genetic operator: crossover
ON GPU: Apply genetic operator: mutation
ON GPU: Apply replacement (new generation created)
end while
Report on results
Clean-up (de-allocate memory on GPU/CPU; release device)
  
```

GPU, graphics processing unit; CPU, central processing unit.

To achieve data parallelism and make use of the CUDA thread hierarchy, we mapped individual genes onto a two-dimensional grid. Using the representation shown in Equation 9, assuming a population of 400 individuals and 18 stocks:

$$\begin{aligned}
 Individual_1 &= \{1,2,4,5,7\dots,6\} \\
 Individual_2 &= \{9,2,1,1,1\dots,2\} \\
 Individual_3 &= \{3,1,3,4,6\dots,2\} \\
 &\vdots \\
 Individual_{400} &= \{8,1,9,8,7\dots,3\}
 \end{aligned}$$

would be mapped to grid cells, as illustrated in Figure 1. The data grid cells are mapped to threads, where each thread executes a kernel processing the data cell at the respective *xy*-coordinate.

Given the hardware used in this investigation (see Table 1), Table 2 outlines the restrictions on the permissible stock universe and population sizes imposed by the chosen mapping of individual genes to threads. A thread block dimension of 32 is chosen for larger problems, because this number ensures that the permissible population size is larger than the number of stocks to cluster.

We note that the efficiency of the algorithm may be compromised near the physical limits outlined in Table 2, as the CUDA memory hierarchy would force threads to access high-latency global memory banks more often. However, for the particular domain problem we are considering here, the Johannesburg Stock Exchange consists of about 400 listed companies on its main board, which represents an upper limit on the number of stocks of interest for local cluster analysis. This is well within the physical limits of the algorithm, while still providing scope to extend the application to multiple markets.

The details on the full implementation, as well as specific choices regarding initialisation, block sizes and threads per block, are given in Cieslakiewicz<sup>13</sup>.

	Individual 1	Individual 2	Individual 3	Individual 4	Individual 5	...	Individual 400
	<b>GRID</b>						
	Block (0,0)	Block (1,0)	Block (2,0)	Block (3,0)	Block (4,0)	...	Block (399,0)
<b>Stock 1</b>	<b>1</b>	<b>9</b>	<b>3</b>	<b>2</b>	<b>5</b>		<b>8</b>
	Block (0,1)	Block (1,1)	Block (2,1)	Block (3,1)	Block (4,1)	...	Block (399,1)
<b>Stock 2</b>	<b>2</b>	<b>2</b>	<b>1</b>	<b>1</b>	<b>7</b>		<b>1</b>
	Block (0,2)	Block (1,2)	Block(2,2)	Block (3,2)	Block (4,2)	...	Block (399,2)
<b>Stock 3</b>	<b>4</b>	<b>1</b>	<b>3</b>	<b>3</b>	<b>3</b>		<b>9</b>
	Block (0,3)	Block (1, 3)	Block (2,3)	Block (3,3)	Block (4,3)	...	Block (399,3)
<b>Stock 4</b>	<b>5</b>	<b>1</b>	<b>4</b>	<b>3</b>	<b>4</b>		<b>8</b>
	Block (0,4)	Block (1,4)	Block (2,4)	Block (3,4)	Block (4,4)	...	Block (399,4)
<b>Stock 5</b>	<b>7</b>	<b>1</b>	<b>6</b>	<b>7</b>	<b>8</b>		<b>7</b>
<b>:</b>	<b>:</b>	<b>:</b>	<b>:</b>	<b>:</b>	<b>:</b>		<b>:</b>
	Block (0,17)	Block (1,17)	Block (2,17)	Block (3,17)	Block (4,17)	...	Block (399,17)
<b>Stock 18</b>	<b>6</b>	<b>2</b>	<b>2</b>	<b>1</b>	<b>2</b>		<b>3</b>

**Figure 1:** Mapping of individuals onto the Compute Unified Device Architecture (CUDA) thread hierarchy.



**Table 2:** Restrictions on the number of stocks and population size

Graphics card	Compute capability	Number of streaming multiprocessors	Maximum threads / thread block	Thread block dimension	Maximum thread blocks / multiprocessor	Maximum number of stocks	Maximum population size
Nvidia GTX Titan Black	3.5	15	1024	32	16	3840	17 472
Nvidia Tesla C2050	2.0	14	1024	32	8	3584	18 720

For the Tesla card, the maximum number of stocks =  $(14) * (1024/32) * 8 = 3584$  and maximum population size =  $(65\ 535 / (3584/32)) * 32 = 18\ 720$ .

### Key implementation challenges

A key challenge in CUDA programming is adapting to the single-program multiple-data (SPMD) paradigm, where multiple instances of a single program use unique offsets to manipulate portions of a block of data.<sup>33</sup> This architecture suits data parallelism, whereas task parallelism requires a special effort. In addition, because each warp (a group of 32 threads) is executed on a single SPMD processor, divergent threads in a warp can severely impact performance. In order to exploit all processing elements in the multi-processor, a single instruction is used to process data from each thread. However, if one thread needs to execute different instructions as a result of a conditional divergence, all other threads must effectively wait until the divergent thread re-joins them. Thus, divergence forces sequential thread execution, negating a large benefit provided by SPMD processing.

The CUDA memory hierarchy contains numerous shared memory banks which act as a common data cache for threads in a thread block. In order to achieve full throughput, each thread must access a distinct bank and avoid bank conflicts, which would result in additional memory requests and reduce efficiency. In our implementation, bank conflicts were avoided by using padding, in which shared memory is padded with an extra element such that neighbouring elements are stored in different banks.<sup>34</sup>

CUDA provides a simple and efficient mechanism for thread synchronisation within a thread block via the `__syncthreads()` barrier function; however, inter-block communication is not directly supported during the execution of a kernel. Given that the genetic operators can be applied only once the entire population fitness is calculated, it is necessary to synchronise thread blocks assigned to the fitness computation operation. We implemented the CPU implicit synchronisation scheme.<sup>35,36</sup> As kernel launches are asynchronous, successive kernel launches are pipelined and thus the executions are implicitly synchronised with the previous launch, with the exception of the first kernel launch. Given the latency incurred on calls between the CPU and GPU, and the consequent drag on performance, GPU synchronisation schemes were explored which achieve the required inter-block communication. In particular, GPU simple synchronisation, GPU tree-based synchronisation and GPU lock-free synchronisation were considered.<sup>35</sup>

Ultimately, the GPU synchronisation schemes were too restrictive for our particular problem, because the number of thread blocks would have an upper bound equal to the number of SMs on the GPU card. If the number of thread blocks is larger than the number of SMs on the card, execution may deadlock. This deadlock could be caused by the warp scheduling behaviour of the GPU, whereby active thread blocks resident on a SM may remain in a busy waiting state, waiting for unscheduled thread blocks to reach the synchronisation point. While this scheme may be more efficient for smaller problems, we chose the CPU synchronisation scheme in the interest of relative scalability.

### Data pre-processing

To generate the  $N$ -stock correlation matrices to demonstrate the viability of the algorithm on real-world test data, data correlations were computed on data, where missing data was addressed using zero-order hold interpolation.<sup>37</sup> The market mode was removed using the

method suggested by Giada and Marsili<sup>17</sup> using a recursive averaging algorithm. A covariance matrix was then computed using an iterative online exponentially weighted moving average filter with a default forgetting factor of  $\lambda=0.98$ . The correlation matrix was computed from the covariance matrix and was cleaned using random matrix theory methods. In particular, Gaussian noise effects were reduced by eliminating eigenvalues in the Wishart range in a trace-preserving manner.<sup>37</sup> This process enhanced the clusters and improved the stability of estimated sequence of correlation matrices.

### Data post-processing

Computed cluster configurations are read from the CUDA output flat file. Successively, an adjacency matrix is constructed by using data values from the correlation matrix in conjunction with computed cluster configuration of the respective data set. The adjacency matrix is then used to construct a disjoint set of minimal spanning trees (MSTs), with each tree capturing the inter-connectedness of each cluster. Each MST exhibits  $n_s - 1$  edges, connecting the  $n_s$  stocks of the cluster in such a manner that the sum of the weights of the edges is a minimum. Kruskal's algorithm was used to generate the MSTs, which depict the linkages between highly correlated stocks, providing a graphical visualisation of the resultant set of disjoint clusters.<sup>38</sup>

## Data and results

### Data

In this investigation we used two sets of data: the training set and the test set. The training set consisted of a simulated time series of 40 stocks which exhibit known distinct, disjoint clusters. The recovery of these induced clusters was used to tune the PGA parameters. The test set consisted of actual stock quoted midprice ticks aggregated into 3-min bars from 28 September 2012 to 10 October 2012, viz. approximately 1800 data points for each stock. Stocks chosen represent the 18 most liquid stocks on the Johannesburg Stock Exchange for that period, according to traded volumes. For both data sets, correlation matrices were constructed from the time series data to serve as inputs for the clustering algorithm. The test set results below show the summary statistics from a set of 1760 correlation matrices of 18 Johannesburg Stock Exchange stocks.

### Results

We show a sample set of results here. Further discussion regarding aspects of the analysis are given in Cieslakiewicz<sup>13</sup>.

### Optimal algorithm settings

Various investigations were undertaken to identify optimal adjoint parameters for the PGA. In each case, the algorithm was successively applied to the training set, with known disjoint clusters. Settings were varied until the rate of convergence was maximised. Once the optimal value for each adjoint parameter had been determined from the training set, the optimal algorithm configuration was deployed on the test set. In further investigations, we will study the effect of various adjoint parameter choices on the rate of convergence and algorithm efficiency for varying stock universe sizes.

The optimal configuration shown in Table 3 for the PGA was deployed on the test set, given a population size of 1000.

**Table 3:** Optimal adjoint parameter values for given training set

Adjoint parameter	Value
Number of generations	400
Crossover probability ( $P_c$ )	0.9
Mutation probability ( $P_m$ )	0.1
Error tolerance	0.00001
Stall generations ( $G_{stall}$ )	50
Elite size	10
Crossover operator	Knowledge-based operator
Mutation operator	Random replacement
Knowledge-based crossover probability	0.9

### Benchmark timing results

Table 4 illustrates the efficiency of the CUDA PGA implementation, compared to the MATLAB serial GA. Direct comparison between the MATLAB serial GA and CUDA PGA may be biased by the fundamental architecture differences of the two platforms. Nevertheless, we immediately observe a significant 10–15 times performance improvement for the test set cluster analysis run. This improvement can be attributed to the utilisation of a parallel computation platform, a novel genetic operator and the algorithm tuning techniques employed. On the GTX platform, the CUDA PGA takes 0.80 s to identify residual clusters inherent in a single correlation matrix of 18 real-world stocks, demonstrating its potential as a near-real-time risk assessment tool. The outperformance of the GTX card is likely explained by the card's relative faster core speed, memory speed, larger memory and memory bandwidth compared to the TESLA card. Although this may justify the use of the more cost-effective GTX card, it is not clear that this performance differential will persist as the size of the stock universe increases, or whether the GTX card preserves solution quality. We note that the scale of the performance improvement over the serial algorithm is not as important as the absolute result of obtaining sub-second computation time. The CUDA PGA thus serves the objective of near-real-time risk assessment, whereby interesting phenomena from emerging stock cluster behaviour can be identified and acted upon to mitigate adverse scenarios. The scalability of these results should be investigated in further research; in particular, the impact of the CUDA memory hierarchy on computation time as global memory accesses increase should be investigated.

In these results we assume correlation matrices are readily available as inputs for the cluster analysis algorithm. Further research to investigate computationally efficient correlation estimation for high-frequency data is a separate problem in the objective of developing a robust and practical near-real-time risk assessment tool.

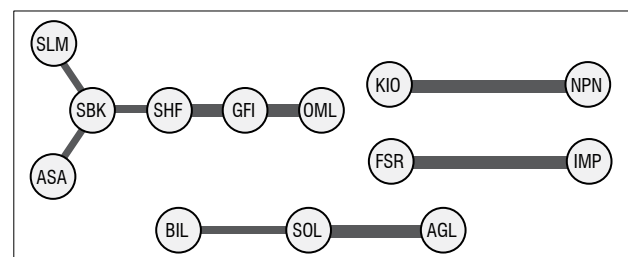
Although the results are promising, it is not clear that the SPMD architecture used by CUDA is well suited for the particular problem considered. The required data dependence across thread blocks

restricts the assignment of population genes to threads and results in a large number of synchronisation calls to ensure consistency of each generation. An MPI island model with distributed fitness computation and controlled migration is perhaps a more well-posed solution<sup>39</sup>; however, it is important to consider the cost of the set-up required to achieve the equivalent speed-up provided by CUDA. This cost should be explored in further research.

### Interpretation of real-world test set results

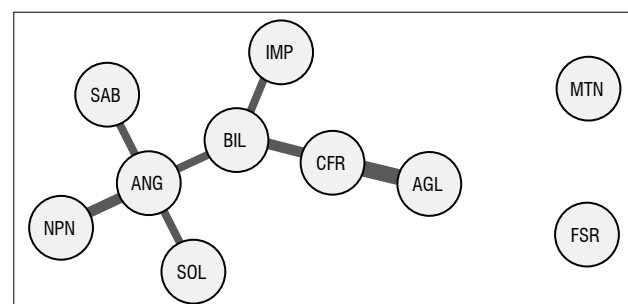
In this section, we illustrate a sample of the resultant cluster configurations which were generated from our model, represented graphically as MSTs.<sup>13,20</sup> This illustration serves as a particular domain application which provides an example of resulting cluster configurations which have meaningful interpretations. The thickness of the edges connecting nodes gives an indication of the strength of the correlation between stocks.

The South African equity market is often characterised by diverging behaviour between financial/industrial stocks and resource stocks and strong coupling with global market trends.



**Figure 2:** Morning trading residual clusters (on 28 September 2012 at 09:03).

In Figure 2, we see four distinct clusters emerge as a result of the early morning trading patterns, just after market open: most notably, a six-node financial/industrial cluster (SLM, SBK, ASA, SHF, GFI, OML) and a three-node resource cluster (BIL, SOL, AGL). At face value, these configurations would be expected; however, we notice that GFI, a gold mining company, appears in the financial cluster and FSR, a banking company, does not appear in the financial cluster. These examples are of short-term decoupling behaviour of individual stocks as a consequence of idiosyncratic factors.



**Figure 3:** Morning trading (after UK market open) residual clusters (on 28 September 2012 at 10:21).

**Table 4:** Benchmark computational speed results

Environment	Framework	Benchmark	Median time (s)	Minimum time (s)	Maximum time (s)
WINDOWS_GTX_CUDA	CUDA-C 5.5	18-stock test set (optimal configuration)	0.80	0.73	3.17
WINDOWS_GTX_MATLAB	Serial	18-stock test set (optimal configuration)	7.77	6.72	13.27
WINDOWS_TESLA_CUDA	CUDA-C 5.5	18-stock test set (optimal configuration)	1.39	1.36	5.51
WINDOWS_TESLA_MATLAB	Serial	18-stock test set (optimal configuration)	15.91	13.41	26.22

Figure 3 illustrates the effect of the UK market open on local trading patterns. We see a clear emergence of a single large cluster, indicating that trading activity by UK investors has a significant impact on the local market. When examining the large single cluster, all of the stocks have either primary or secondary listings in the USA and UK. In particular, SAB and ANG have secondary listings on the London Stock Exchange, whereas BIL and AGL have primary listings on the London Stock Exchange.<sup>40</sup> It is also unusual to see such a strong link (correlation) between AGL, a mining company, and CFR, a luxury goods company. This correlation may be evidence that significant UK trading in these two stocks can cause a short-term elevated correlation, which may not be meaningful or sustainable.

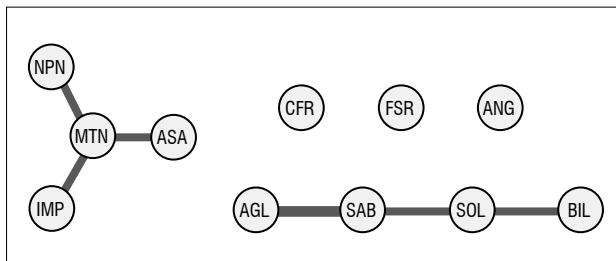


Figure 4: Midday trading residual clusters (on 28 September 2012 at 12:21).

In Figure 4, we consider midday trading patterns. We see that the clustering effect from UK trading has dissipated and multiple disjoint clusters have emerged. CFR has decoupled from AGL in the 2 h after the UK market open, as we might expect. We see a four-node financial/industrial cluster (NPN, MTN, ASA, IMP) and a four-node resource cluster (AGL, SAB, SOL, BIL); IMP, a mining company, appears in the financial/industrial cluster.

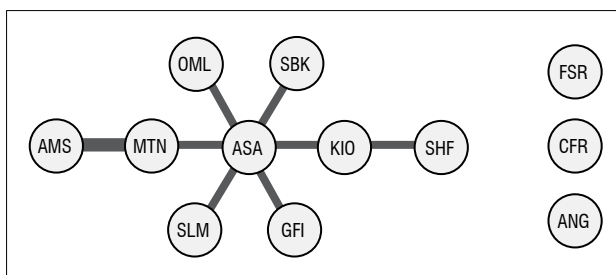


Figure 5: Afternoon trading (after US market open) residual clusters (on 28 September 2012 at 15:33).

Figure 5 illustrates the effect of the US market open on local trading patterns. Similar to what we observed in Figure 3, we see the emergence of a large single cluster, driven by elevated short-term correlations amongst constituent stocks. This observation provides further evidence that significant trading by foreign investors in local stocks can cause a material impact on stock market dynamics.

## Conclusion

In this paper, we have verified that the Giada and Marsili<sup>11</sup> likelihood function is a viable, parallelisable approach for isolating residual clusters in data sets on a GPU platform. Key advantages of this function compared with conventional clustering methods are that: (1) the method is unsupervised and (2) the interpretation of results is transparent in terms of the model.

The implementation of the master-slave PGA showed that efficiency depends on various algorithm settings. The type of mutation operator utilised has a significant effect on the algorithm's efficiency to isolate the optimal solution in the search space, whilst the other adjoint parameter settings primarily impact the convergence rate. According to the benchmark test results, the CUDA PGA implementation runs 10–15 times faster than the serial GA implementation in MATLAB for detecting

clusters in 18-stock real-world correlation matrices. Specifically, when using the Nvidia GTX Titan Black card, clusters are recovered in sub-second speed, demonstrating the efficiency of the algorithm.

Provided intraday correlation matrices can be estimated from high frequency data, this significantly reduced computation time suggests intraday cluster identification can be practical, for near-real-time risk assessment for financial practitioners.

Detecting cluster anomalies and measuring persistence of effects may provide financial practitioners with useful information to support local trading strategies. From the sample results shown, it is clear that intraday financial market evolution is dynamic, reflecting effects which are both exogenous and endogenous. The ability of the clustering algorithm to capture interpretable and meaningful characteristics of the system dynamics, and the generality of its construction, suggests the method can be successful in other domains.

Further investigations will include adjoint parameter tuning and performance scalability for varying stock universe sizes and cluster types, quantifying the variability of solution quality on the GTX architecture as a result of non-ECC memory usage and the investigation of alternative cost-effective parallelisation schemes. Given the SPMD architecture used by CUDA, the required data dependence across thread blocks restricts the assignment of population genes to threads and results in a large number of synchronisation calls to ensure consistency of each generation. An MPI island model with distributed fitness computation and controlled migration is perhaps a more well-posed solution to explore<sup>39</sup>; however, the cost of the set-up required to achieve the equivalent speed-up provided by CUDA should be justified.

## Acknowledgements

This work is based on research supported in part by the National Research Foundation (NRF) of South Africa (grant numbers 87830, 74223 and 70643). The conclusions herein are those of the authors and the NRF accepts no liability in this regard. We thank Margit Haerting and Michelle Kuttel for their feedback on preceding projects culminating in this work.

## Authors' contributions

D.W. and T.G. were responsible for the initial conception of the problem and idea development. T.G. assisted in data provision and cleaning, as well as in implementation of the serial algorithm. D.H. implemented the code, performed simulations, collated results and wrote the manuscript.

## References

- Luque G, Alba E. Parallel genetic algorithms. Berlin: Springer-Verlag; 2011. [http://dx.doi.org/10.1007/978-3-642-22084-5\\_1](http://dx.doi.org/10.1007/978-3-642-22084-5_1)
- Colomi A, Dorigo M, Maffioli F, Maniezzo V, Righini G, Trubian M. Heuristics from nature for hard combinatorial optimization problems. *Int Trans Oper Res*. 1996;3:1–21. <http://dx.doi.org/10.1111/j.1475-3995.1996.tb00032.x>
- Bohm C, Noll R, Plant C, Wackersreuther B. Density-based clustering using graphics processors. In: Proceedings of the 18th ACM Conference on Information and Knowledge Management. New York: ACM; 2009. p. 661–670. <http://dx.doi.org/10.1145/1645953.1646038>
- Brecheisen S, Kriegel HP, Pfeifle M. Parallel density-based clustering of complex objects. In: Ng W-K, Kitsuregawa M, Li J, Chang K, editors. Advances in knowledge discovery and data mining. Proceedings of the 10th Pacific-Asia Conference; 2006 Apr 9–12; Singapore. Berlin: Springer; 2006. p. 179–188. [http://dx.doi.org/10.1007/11731139\\_22](http://dx.doi.org/10.1007/11731139_22)
- Dessel T, Anderson DP, Magdon-Ismael M, Newberg H, Szymanski BK, Varela CA. An analysis of massively distributed evolutionary algorithms. In: Proceedings of 2010 IEEE Congress on Evolutionary Computation; 2010 July 18–23; Barcelona, Spain. IEEE; 2010. p. 1–8. <http://dx.doi.org/10.1109/CEC.2010.5586073>
- Jaimes A, Coello Coello C. MRMOGA: A new parallel multi-objective evolutionary algorithm based on the use of multiple resolutions. *Concurrency Comput Pract Experience*. 2007;19(4):397–441. <http://dx.doi.org/10.1002/cpe.1107>

7. Kromer P, Platos J, Snasel V. Data parallel density-based genetic clustering on the CUDA architecture. *Concurrency Comput Pract Experience*. 2014;26:1097–1112. <http://dx.doi.org/10.1002/cpe.3054>
8. Pospichal P, Jaros J, Schwarz J. Parallel genetic algorithm on the CUDA architecture. In: Di Chio C, Cagnoni S, Cotta C, Ebner M, Ekárt A, Esparcia-Alcazar AI, et al., editors. *Applications of Evolutionary Computation EvoApplications 2010: EvoCOMPLEX, EvoGAMES, EvoASP, EvoINTELLIGENCE, EvoNUM, and EvoSTOC*; 2010 Apr 7–9; Istanbul, Turkey. Berlin: Springer; 2010. p. 442–451. [http://dx.doi.org/10.1007/978-3-642-12239-2\\_46](http://dx.doi.org/10.1007/978-3-642-12239-2_46)
9. Robilliard D, Marion M, Fonlupt C. High performance genetic programming on GPU. In: *Proceedings of the 2009 Workshop on Bio-inspired Algorithms for Distributed Systems*. New York: ACM; 2009. p. 85–94. <http://dx.doi.org/10.1145/1555284.1555299>
10. Tirumalai V, Ricks K, Woodbury K. Using parallelization and hardware concurrency to improve the performance of a genetic algorithm. *Concurrency Comput Pract Experience*. 2007;19:443–462. <http://dx.doi.org/10.1002/cpe.1113>
11. Giada L, Marsili M. Algorithms of maximum likelihood data clustering with applications. *Physica A*. 2002;315:650–664. [http://dx.doi.org/10.1016/S0378-4371\(02\)00974-3](http://dx.doi.org/10.1016/S0378-4371(02)00974-3)
12. Omran M, Salaman A, Engelbrecht A. Dynamic clustering using particle swarm optimization with application in image segmentation. *Pattern Anal Appl*. 2006;8:332–344. <http://dx.doi.org/10.1007/s10044-005-0015-5>
13. Cieslakiewicz D. Unsupervised asset cluster analysis implemented with parallel genetic algorithms on the Nvidia CUDA platform [thesis]. Johannesburg: University of the Witwatersrand; 2014.
14. Everitt B, Landau S, Leese M. *Cluster analysis*. 5th ed. Chichester: Wiley; 2001.
15. Blatt M, Wiseman S, Domany E. Superparamagnetic clustering of data. *Phys Rev Lett*. 1996;76:3251–3254. <http://dx.doi.org/10.1103/PhysRevLett.76.3251>
16. Kullmann L, Kertész J, Mantegna RN. Identification of clusters of companies in stock indices via Potts super-paramagnetic transitions. *Physica A*. 2000;287(3–4):412–419. [http://dx.doi.org/10.1016/S0378-4371\(00\)00380-0](http://dx.doi.org/10.1016/S0378-4371(00)00380-0)
17. Giada L, Marsili M. Data clustering and noise undressing of correlation matrices. *Phys Rev E*. 2001;63(6), Art. #061101, 8 pages. <http://dx.doi.org/10.1103/PhysRevE.63.061101>
18. Noh JD. A model for correlations in stock markets. *Phys Rev E*. 2000;61(5):5981–5982. <http://dx.doi.org/10.1103/PhysRevE.61.5981>
19. Sornette D. *Critical phenomena in natural sciences: Chaos, fractals, selforganization and disorder: Concepts and tools*. Berlin: Springer; 2000. <http://dx.doi.org/10.1007/978-3-662-04174-1>
20. Mbambiso B. *Dissecting the South African equity markets into sectors and states* [thesis]. Cape Town: University of Cape Town; 2009.
21. Gebbie T, Wilcox D, Mbambiso B. Spin, stochastic factor models, and a GA. PDF presented at: Southern African Finance Association conference; 2010. [http://www.academia.edu/2499934/Spin\\_Stochastic\\_Factor\\_Models\\_and\\_a\\_GA](http://www.academia.edu/2499934/Spin_Stochastic_Factor_Models_and_a_GA).
22. Sivanandam S, Deepa S. *Introduction to genetic algorithms*. Berlin: Springer Science & Business Media; 2010.
23. Baker J. Reducing bias and inefficiency in the selection algorithm. In: Grefenstette JJ, editor. *Proceedings of the Second International Conference on Genetic Algorithms and their Application*. Hillsdale, NJ: L. Erlbaum Associates Inc.; 1987. p. 14–21.
24. Coley D. *An introduction to genetic algorithms for scientists and engineers*. Singapore: World Scientific Publishing; 1999. <http://dx.doi.org/10.1142/3904>
25. Ismail M. Parallel genetic algorithms (PGAs): Master-slave paradigm approach using MPI. *IEEE e-Tech*. 2004;31:83–87. <http://dx.doi.org/10.1109/ETECH.2004.1353848>
26. Digalakis J, Margaritis K. Parallel evolutionary algorithms on message-parsing clusters: Working paper [document on the Internet]. c2003 [cited 2015 May 12]. Available from: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.3.2539&rep=rep1&type=pdf>
27. Zhang J, Liu W, Liu G. Parallel genetic algorithm based on the MPI environment. *Telkomnika*. 2012;10:1708–1715. <http://dx.doi.org/10.11591/telkomnika.v10i7.1566>
28. Nan L, Pengdong G, Yongquan L, Wenhua Y. The implementation and comparison of two kinds of parallel genetic algorithm using Matlab. In: *Ninth International Symposium on Distributed Computing and Applications to Business, Engineering and Science (DCABES)*; 2010 Aug. 10–12; Hong Kong, China. IEEE; 2010. p. 13–17. <http://dx.doi.org/10.1109/DCABES.2010.9>
29. Nvidia. *Nvidia CUDA C programming guide*. Santa Clara, CA: Nvidia Corporation; 2011.
30. Nvidia. *CUDA dynamic parallelism programming guide*. Santa Clara, CA: Nvidia Corporation; 2012.
31. ACT HPC. GTX vs Tesla [blog on the Internet]. c2014 [cited 2014 Sep 25]. Available from: <http://www.advancedclustering.com/hpc-cluster-blog-gtx-vs-tesla>
32. Zhang S, He Z. Implementation of parallel genetic algorithm based on CUDA. In: Cai Z, Li Z, Kang Z, Liu Y, editors. *Advances in Computation and Intelligence. Proceedings of the 4th International Symposium on Advances in Computation and Intelligence*; 2009 Oct 23–25; Huangshi, China. Berlin: Springer; 2009. p. 24–30. [http://dx.doi.org/10.1007/978-3-642-04843-2\\_4](http://dx.doi.org/10.1007/978-3-642-04843-2_4)
33. Darema F. SPMD model: Past, present and future. In: Cotronis Y, Dongarra J. *Recent advances in parallel virtual machine and message passing interface: Proceedings of the 8th European PVM/MPI Users' Group Meeting*; 2001 Sep 23–26; Santorini, Greece. Berlin: Springer; 2001. p. 1. [http://dx.doi.org/10.1007/3-540-45417-9\\_1](http://dx.doi.org/10.1007/3-540-45417-9_1)
34. Brodtkorb A, Hagen T, Saetra M. Graphics processing unit (GPU) programming strategies and trends in GPU computing. *J Parallel Distr Com*. 2012;73(1):4–13. <http://dx.doi.org/10.1016/j.jpdc.2012.04.003>
35. Xiao S, Feng W. Inter-block GPU communication via fast barrier synchronization. In: *2010 IEEE International Symposium on Parallel & Distributed Processing (IPDPS)*; 2010 Apr 19–23; Atlanta, GA, USA. IEEE; 2010. p. 1–12. <http://dx.doi.org/10.1109/IPDPS.2010.5470477>
36. Nvidia. *CUDA C best practices guide*. Santa Clara, CA: Nvidia Corporation; 2012.
37. Wilcox D, Gebbie T. An analysis of cross-correlations in South African market data. *Physica A*. 2007;375:584–598. <http://dx.doi.org/10.1016/j.physa.2006.10.030>
38. Kruskal J. On the shortest spanning subtree of a graph and the traveling salesman problem. *Proc Amer Math Soc*. 1956;7:48–50. <http://dx.doi.org/10.1090/S0002-9939-1956-0078686-7>
39. Whitley D, Rana S, Heckendorn R. The island model genetic algorithm: On separability, population size and convergence. *J Comput Inform Technol*. 1999;7:33–47.
40. JSE. *Companies and financial instruments* [homepage on the Internet]. c2013 [cited 2013 Sep 25]. Available from: <https://www.jse.co.za/current-companies/companies-and-financial-instruments>





# Publishing patterns at the Cape Peninsula University of Technology

## AUTHORS:

Elisha Chiware<sup>1</sup>

Lara Skelly<sup>2</sup>

## AFFILIATIONS:

<sup>1</sup>CPUT Libraries, Cape Peninsula University of Technology, Cape Town, South Africa

<sup>2</sup>Stellenbosch University Business School, Cape Town, South Africa

## CORRESPONDENCE TO:

Lara Skelly

## EMAIL:

Lara.Skelly@gmail.com

## POSTAL ADDRESS:

Cape Peninsula University of Technology, PO Box 652, Cape Town 8000, South Africa

## DATES:

Received: 01 July 2014

Revised: 30 Oct. 2014

Accepted: 17 Feb. 2015

## KEYWORDS:

bibliometrics; journal impact; citations; research visibility; South Africa

## HOW TO CITE:

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. <http://dx.doi.org/10.17159/sajs.2016/20140220>

We report on the publishing and collaborative patterns of researchers in the applied sciences and engineering disciplines at a technological university in South Africa, Cape Peninsula University of Technology. The study focused on works published over a 10-year period (2005–2014) and was based on citations of peer-reviewed literature from Scopus. The results showed a steady growth in research outputs in science and technology at the university. There has also been a growing trend of international collaboration in certain disciplines. This scoping study serves as a benchmark for similar studies.

## Introduction

The aim of this paper was to present publishing patterns of science researchers in the applied sciences and engineering disciplines at Cape Peninsula University of Technology (CPUT). While the university is not yet classified as a research-intensive institution, its research outputs have been growing since its inception in 2005, which was the result of the merger of two technical colleges. Historically, research at South African technical colleges was not a priority as the focus was more on skills training and development. But since 1994, and the subsequent mergers of various institutions in the mid-2000s, the growth and development of research activities in these former technical colleges has been growing, partly as a result of the new mandates of such institutions.

The rapid growth of research output at CPUT is a result of sustained efforts to grow the institutional research landscape in response to government's reorientation of science and technology through: research, technology and innovation policies, strong research governance systems, increased research funding, recruitment of highly skilled staff, research infrastructure development and building synergies within the university's structures to provide the necessary support to researchers. The development of the research technology, innovation and partnerships blueprint,<sup>1</sup> establishment and appointment of several top-rated researchers as research chairs, participation in various international research collaborative programmes as well as mentoring and supporting young researchers and growing the number of postgraduate students and postdoctoral positions, have all contributed towards a rich research environment within the institution. Since the inception of CPUT, research has been growing in many areas, most notably in Applied Sciences and Engineering. The focus now is to steer the university's research activities towards multi-disciplinary research focus areas. These focus areas are:

- Bio-economy and biotechnology
- Space science and technology
- Energy
- Climate change and environment
- Human and social dynamics, including issues related to service delivery
- Economic growth and international competitiveness
- Design for sustainability

There is growing interest in studies on bibliometric analysis, journal impact and collaborative research patterns in African universities and research institutions as a result of various demands and the desire to align with global trends. Given the importance of the global rankings of universities based on research outputs and the demand by funders regarding the impact of research on development, there is a need to evaluate the relevance and impact of the research outputs in African institutions. It is also important to determine the overall impact they are making towards economic and social development on the continent. Such studies will contribute towards research and development projects as well as the uptake of research outputs, especially in areas of greatest need such as poverty alleviation, health care, increased food production, appropriate technology uptake and the general improvement of the lives of people in African countries. The studies on African scientific publishing patterns point to a preference by established and renowned researchers to publish in leading global journals as well those that are nationally accredited, as is the case in South Africa where the Department of Higher Education and Training offers subsidies based on publishing in certain accredited journals and recognised lists.<sup>2</sup>

It has become a national norm for researchers in tertiary education institutions to publish and publishing is part of the performance management and promotion criteria of several institutions on the continent. Other factors contributing to increased research activities and outputs include increased funding, significant policy changes, improved research infrastructure (both human and physical), information and communications technology resources, open, free and low cost access to peer reviewed literature and research capacity building training,<sup>3</sup> and increased collaborative opportunities with colleagues in well-resourced institutions abroad. According to Li et al.<sup>4</sup>, maintaining high levels of research productivity is essential for researchers' careers if they are to impact the knowledge community through citations and establish reputations as scholars. In Latin America, Holmgren and Schnitzer<sup>5</sup> attributed the increased scientific productivity to the strong scientific developments in many countries in the region during the 1990s. The contribution also came from increased funding to the most productive scientists

through international cooperation with scientists from Europe and the United States and collaboration among scientists within Latin America.

Journal rankings remain an important factor in the lives of researchers and academic communities. Various methods are applied to determine both author and journal productivity and these methods range from citation studies to perception and market test studies. The most common tools used are citation abstracts and indexes, opinion surveys and the analysis of journal holdings.<sup>6</sup> Pouris<sup>7</sup> points out that scientometric studies, in general, and collaboration studies, in particular, are in an embryonic stage. Furthermore, he emphasises that even South Africa, which is a major producer of research publications on the continent, produces few publications in scientometrics.

## Literature review

### *Publishing trends in developing countries*

Researchers in any field tend to follow established publishing patterns because they want to gain as much impact as is possible. This approach is driven by institutional demands, as the impact factor is often one of the few measures that universities look at when considering the quality of a researcher's work. The importance given to established and respected journal titles also supports the tendency to follow established publishing patterns. Several studies on publishing patterns of researchers in developing countries show that the majority prefer to publish in foreign journals.<sup>8</sup> The visibility of the overall scientific research output from developing countries in general, and African countries in particular, is still very low in global terms although the recent past has shown an increased growth. Many contemporary commentators and analysts argue that scientific research in Africa is lagging far behind other regions in the world and is in dire need of large investments in order to catch up with other developing regions.<sup>9</sup> Competition with established researchers from developed countries is high for entry into established journal titles. In addition, in most cases, the fields of research by African researchers are not viewed as important for global audiences and hence their marginalisation in leading journals. The alternative is often national journals which are poorly represented in international citation tools, resulting in poor visibility of the researchers, their outputs<sup>10</sup> and impact. There is an increasing trend to publish in open access journals and again, some of these suffer the same fate as titles published in developing countries. In terms of research output, however, there are high outputs, notably from rising giants like China, India and Brazil, as well as some of the major countries in southeast Asia and in Africa, where Egypt, Nigeria and South Africa are notably the continent's publishing powerhouses, contributing about 80% of the continent's scientific output.<sup>11</sup>

Although notable achievements have been made in some developing regions, Latin America and Africa included, gaps still exist<sup>5</sup> and there is still reason to push for greater achievement.

### *Scientific research collaboration*

Scientific research is becoming increasingly global, interdisciplinary and collaborative in nature<sup>12,13</sup> and critical to scientific practice.<sup>14</sup> It is widely assumed that collaboration in research is a positive development and that it should be encouraged.<sup>15</sup> In addition to high funding levels in the fields of science and technology, improved and changing communication channels, as well as the mobility of researchers, have all enabled collaboration at a global level.<sup>16</sup> In recent years, increased collaboration has been viewed as good practice that improves institutional performance and global rankings of universities. Other benefits, such as attracting good researchers and more funding, are all well documented.<sup>17</sup>

Ynalez and Shrum<sup>14</sup> see scientific collaboration as emanating from the social structure of relations among knowledge producers and successful research projects, translating into new resources that facilitate knowledge generation or the attainment of specified objectives.

One of the most common and well established forms of scientific research collaboration is co-authorship, which can be effectively traced and analysed through bibliometric studies.<sup>16</sup> Co-authorship increasingly creates networks that reveal certain features of the participating

academic communities.<sup>18</sup> According to Bidault and Hildebrand<sup>19</sup>, these arrangements have become prevalent in research and development activities where organisations are seeking partnerships to complement their own technological capabilities. Co-authorships have several advantages for researchers in both developed and developing environments. Li et al.<sup>4</sup> note that co-authoring with prolific scholars helps researchers develop centralities and in turn, generate higher numbers of citations. However, Wagner and Leydesdorff<sup>20</sup> indicate that there is still no full explanation for the rapid growth in scientific international research collaboration. In trying to find answers to this, they tested the hypothesis that international collaboration is a self-organising network and that the growth of co-authorships can be explained based on the organising principle of preferential attachment, although the attachment mechanism deviates from an ideal power law.

Collaboration by African researchers can be viewed from different perspectives. According to Pouris and Ho<sup>21</sup>, African researchers' collaboration with international partners has grown dramatically by 66% over a recent five-year period. The analysis by these authors shows a dominance of research areas in the medical and natural resources fields. The main drivers of these collaborations are the availability of resources and interests outside the continent i.e. international imperatives that often favour group rather than individual research. Onyancha and Maluleka<sup>22</sup> have shown in a study on collaborative research in sub-Saharan Africa that knowledge production through collaboration among regional countries is minimal and contributes only a small percentage when compared to collaboration with countries in other parts of the world.

Research on collaboration among the Southern African Development Community's (SADC) fifteen countries shows that only 3% of outputs between 2005 and 2008 were jointly authored by researchers from two or more SADC countries (intra-regional collaboration), and only 5% of the papers were jointly authored by researchers from African countries outside the SADC region.<sup>23</sup> This is not an ideal trend as there are many opportunities for more cooperation given the relative levels of development within the region, as well as better infrastructure and funding compared with the rest of the continent. According to Sooryamoorthy<sup>24</sup>, 'the existing networks with SADC and NEPAD (New Partnership for African development) have turned out to be unsuitable for want of resources, preventing effective collaboration'. This could also be a sign of a lack of a regional strategy to approach research and development in a more integrated manner within the development community. In addition, the little intra-regional collaborative research that is taking place is largely funded by external donors, again showing that external forces are the key drivers of collaboration among African scholars.

### *Collaboration and publications in South Africa*

According to Baker<sup>2</sup>, the South African scholarly publishing landscape is required to tackle the realities of global science publishing, maintain its existing strengths and have the foresight, based on practical experience, to create a future of safety and stability for the country's best scholarly journals in an unsafe and ever competitive world. Interdisciplinary research in and amongst South African research and academic institutions requires in-depth analysis. Sooryamoorthy<sup>25</sup> points to the growing trend of collaborative research among South African scientists and indicates that this collaboration is mostly international, compared to domestic ventures and that publication seems to be a decisive factor. While institutions might want to safeguard their territories, there is need at all levels to encourage inter-institutional collaborations. In some scientific and technological disciplines, collaboration is active and is producing good results; however, additional areas need to be identified to make this more meaningful.

Leading researchers in South Africa and other leading African research countries like Egypt, Algeria, Morocco and Nigeria, have a tendency to help the introduction of new knowledge into their countries through collaboration, because they are better connected to external sources of knowledge than less competent researchers.<sup>26</sup> The limitation of resources is a feature in the making of such decisions as are the international partners who prefer to work with well-established scholars as a means of advancing the knowledge frontier.

South Africa, the highest producer of publications on the continent, has a collaboration rate of 53% – far higher than fellow BRICS countries which reflect lower rates of collaboration: 25% for Brazil, 20% for India and 23% for China.<sup>27</sup> Barnard et al.<sup>26</sup> point out that this trend runs the risk of world-class researchers completely disconnecting themselves from the rest of the South African research community. The conclusion is however, that local leading researchers are not completely disconnected from upcoming researchers and do in fact work with networks of local partners. Dell<sup>27</sup> points out that 'South Africa's high international collaboration rate persist in spite of the fact that national university funding systems acts as a disincentive to inter-institutional collaboration in the sense that collaborating institutions are required to share the government subsidy that rewards staff members who publish.'

Overall collaboration has had a positive impact on the productivity of South African scientists<sup>28</sup> and scientists prefer international to domestic partnerships. The roots of collaboration date into the past and with time, the country has become a regional hub for collaboration, given its relative strength in higher education, better funding schemes and policy frameworks, well developed research and ICT infrastructures and an attraction to leading African and global scholars. These trends are likely to continue into the future.

Improved visibility is a motive for collaboration.<sup>20</sup> It would stand to reason that examining co-authorship patterns would give one an approximation for visibility (as was undertaken by Sooryamoorthy<sup>29</sup>). While the visibility of African research remains low,<sup>9</sup> the visibility of South African research would have been even lower, had South African researchers not engaged in international collaboration efforts.<sup>24</sup>

## Methods

As the South African academic and research landscape moves towards more internationalisation, it is important to begin to analyse the existing collaboration through various research approaches. Pouris and Ho<sup>21</sup> point out that 'scientific co-authorship of African researchers has become a fashionable topic in recent scientometric literature. Researchers are investigating the effects, modes, dynamics and motives of collaboration...' Bibliometric studies, evaluation of author networks and journal impact, are all areas that need a sustained level of evaluation in order to contribute to the global body of knowledge in these fields.

There is a multitude of perspectives that one could take when exploring research patterns. In this study, the perspectives focused on three areas: general growth, visibility and collaboration. Other perspectives, such as research uptake, can be explored in further studies. This study considered the following:

- Has CPUT research output exceeded the growth rate of other Universities of Technology (UoTs) in South Africa? Has it exceeded the growth rate of the other 22 public universities in the country?
- How visible is CPUT research output when compared with the other UoTs and public universities in general?
- What are the patterns of collaboration with regard to degree and type of CPUT research?

### Method and data

All data were drawn from Scopus on 3–4 February 2015. The data covered the ten-year period from 2005 to 2014. As Scopus has an export limitation of 2000 records, data were drawn for each institution individually. Sol Plaatje University and the University of Mpumalanga had no records for that time period and were excluded. This left 23 institutions in the sample.

The subject areas were limited to those under the engineering and applied science disciplines and were consistently applied to ensure comparability of the results. The following subject areas were included:

- Agricultural and Biological Sciences
- Chemical Engineering
- Chemistry

- Computer Science
- Materials Science
- Energy
- Engineering
- Environmental Science
- Materials Science
- Mathematics
- Physics and Astronomy
- Earth and Planetary Sciences

The classification of articles into these subject areas was done by Scopus. It is recognised that the selection of these subject areas is subjective and other needs might yield a different selection.

Once the records were found in Scopus, they were exported in comma-delimited format. This allowed for the manipulation of records in Microsoft Excel and VOSViewer.<sup>30</sup> VOSViewer was used to glean the number of authors, the number of unique journal titles and the collaboration details. These were collated in Microsoft Excel.

In all cases, 2005 was used as a base year and growth of any variable reflected the change from the base year. As there were three levels of analysis – the institution, UoTs as a group, and all the public universities in South Africa – using the growth rate for comparison was easier that using the raw figures.

### Limitations

This study restricted itself to data available in Scopus. Not all research output was available in Scopus, therefore a larger pool might have yielded more reliable results. However, Scopus is the largest citation database and represents the best data available. Coverage by Scopus of Applied Science and Engineering, the disciplines in question was considered to be better than coverage of the social sciences. By limiting the study to Applied Science and Engineering, the reliability of the results was improved.

Scopus restricts the export of records to an amount of 2000. Without this restriction, the study could have been expanded to include all of African institutions, and possibly institutions on other continents as well.

## Results

As indicated by the research questions, the results were discussed in three groupings: growth, visibility and collaboration. For the first two groupings, comparisons were made between CPUT, UoTs and South African public universities. For the third, collaboration, only CPUT was discussed. Motivations for this restriction are given in the section on collaboration.

### Growth rate

Figure 1 shows how the number of articles has grown over the ten-year period. From 2008, CPUT has grown its output faster than the national average for public universities, and faster than the UoT average. The dip in 2014 shows that South African research output is not exempt from the publication lag experienced by others. The growth of the number of articles indicates that CPUT has responded well to incentives to increase research output, and has embraced the challenge of joining the research community in South Africa.

Considering the publishing cycle, the initial low output from 2005 to 2008 is not surprising. During this time, inexperienced researchers would be going through a period of revising and resubmitting articles. It would appear that after an initial three years, researchers gained a certain level of experience which allowed publication output to increase rapidly.

According to Kahn, the increase in publications can be attributed in part to the increase in publication subsidy by Department of Higher Education and Training, the increase in South African journals indexed in Web of

Science and a shift in focus to fields with higher publication propensity.<sup>31</sup> Pouris<sup>32</sup> points out that the primary incentive fuelling the recent growth is 'the new funding formula in the country which subsidises the universities by more than ZAR100 000 for each publication; the increase in the number of journals indexed in the ISI Thomson Reuters database and the incorporation of social sciences at the NRF'. All these factors contributed towards the growth of research outputs in South Africa.<sup>32</sup>

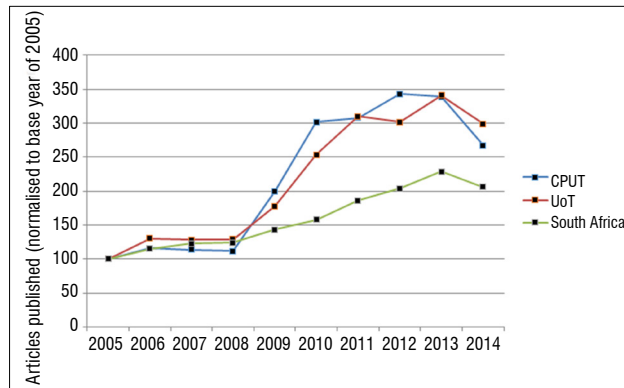


Figure 1: Growth in research output over a ten-year period.

### Visibility

Visibility can be considered within the context of three variables: citations, the number of authors and the number of unique journal titles. All three variables show that there is an increasing visibility of CPUT research which conforms or surpasses the rate shown by South African public institutions.

### Citations

Citation counting is the most common method of measuring visibility. The shortcomings of this method have been thoroughly explored by others<sup>33</sup> and will not be repeated here. Regardless of such criticism, it remains a simple metric for showing the use of research by other researchers.

The growth rate shown in Figure 2 has been normalised for the year of publication. On the whole, the growth rate of citations of CPUT research has been greater than that of both UoTs and South African public institutions in general. The decrease in recent years could be as a result of the citation lag, despite the attempt to normalise for it. A study covering a longer period would be able to reveal if this is the case.

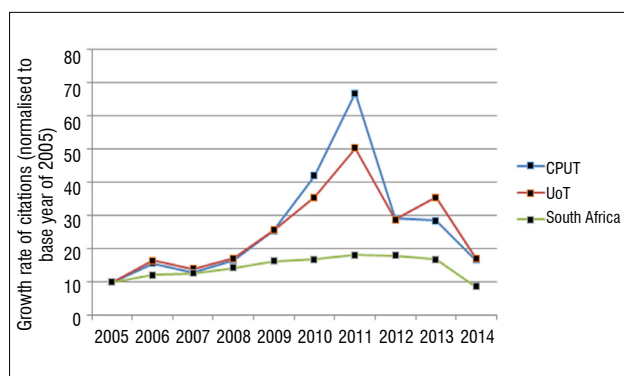


Figure 2: Growth in rate of citations over a 10-year period

### Authors

Every researcher has his or her own community, network or audience. By involving more participants in the authoring of research articles, an institution can increase the visibility of its research. Increasing the number of authors could be done through the encouragement of new researchers to publish or by involving authors outside the institution through co-authorship. Figure 3 does not distinguish between these two

methods, because from the perspective of growing one's audience, they should not differ.

The growth rate of CPUT authors (and co-authors) has been increasing steadily at a rate similar to that of other UoTs and South African institutions in general. In 2011, however, there was a marked increase by the latter, while the UoTs, CPUT included, did not follow suit. Whether this is the start of a levelling off, or simply a short slump, is uncertain. What is clear though is that CPUT has increased the number of authors (and co-authors) by more than three times, which is a significant amount.

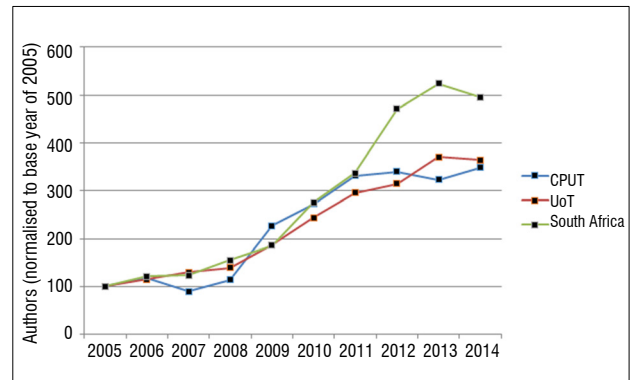


Figure 3: Growth in the number of authors over a 10-year period.

### Journal titles

Another way of increasing one's audience is to publish in more journals, as each journal has (assumedly) its own audience. Figure 4 shows how the number of unique journal titles has grown for CPUT, UoTs and South African public institutions. CPUT has been successful in finding new avenues of publication and the growth rate of unique titles CPUT exceeds other UoTs. Other South African researchers have been growing their audience in the same way, but not at the rate of the UoTs.

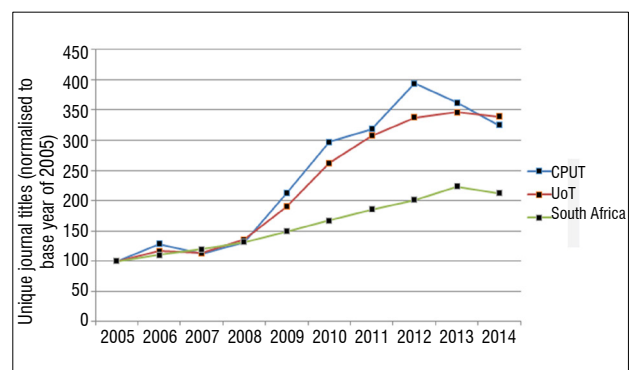


Figure 4: Growth in unique journal titles over a 10-year period.

### Collaboration

The final research question was based on the distinction between degree and type of collaboration made by Sooryamoorthy.<sup>25</sup> Degree is the number of collaboration activities, whereas type refers what geographical boundaries the collaboration transverses. For this research question, collaboration was narrowly considered as co-authorship.

Figure 5 displays how the co-authoring relations have changed over the ten-year period of interest. Co-authorships between researchers at CPUT grew dramatically, showing how the internal research community is bonding. Co-authorships with partners in the Western Cape were noticeably more than co-authorships with other South African partners. There was also more engagement with non-African countries than with other African countries. Figure 5 shows that CPUT researchers are not exploiting geographical advantages by partnering with close neighbours. This could indicate a trend towards e-research.



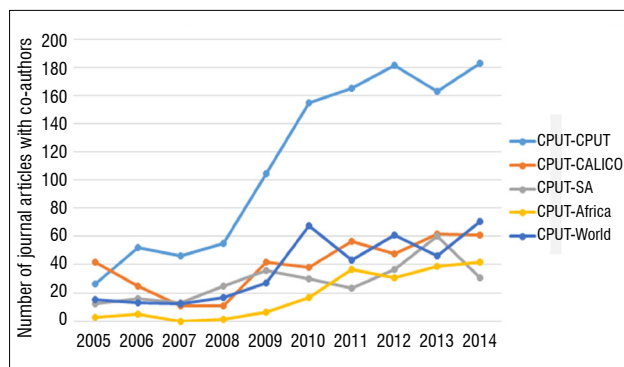


Figure 5: Growth in CPUT collaboration over a 10-year period.

## Conclusions

This paper contributes to the literature on publishing patterns of science and technology in developing countries and represents one of the few scientometric studies on a South African institution. The objectives of the paper were to establish the publishing patterns of researchers in the faculties of Applied Science and Engineering at CPUT and the results show a steady growth in research outputs in science and technology at the university. It is clear that researchers in science and technology disciplines at CPUT have taken up the mandate to contribute to scholarly literature since the inception of the institution in 2005. It also shows a continued trend of international collaboration, not only with developed countries, but with African partners too. The increased visibility of the knowledge produced at CPUT creates a virtuous loop of improving the impact of CPUT research. Evidence of this will be seen in citations.

The use of bibliometric data, as provided by Scopus, is just one of the many ways in which researching patterns can be studied. As a UoT, CPUT is also engaged in creating commercially viable knowledge and contributing positively to the community outside of the university. Bibliometric measures, while useful to a degree, do not paint a full picture on the knowledge production and uptake by the institution.

Looking to the future, CPUT is investigating how to continue to increase the experience of researchers and push past the plateau evident in recent years. Engaging in multidisciplinary research is one method as it draws on everyone's strengths. Increased collaboration, particularly with other developing countries could aid, not only CPUT, but their partners as well. The focus on multidisciplinary research areas, growth of specialised research units, appointment of more Research Chairs, unlocking the potential of staff and students, emphasis on research uptake activities as well as the maintenance of excellence in research will all contribute to the growth and development of research outputs within this young university. These measures will work towards increasing the quantity of publications without sacrificing the quality of the research produced.

This paper focused only on three perspectives of research patterns: growth, visibility and collaboration. Other perspectives, such as research uptake, can be explored to supplement the results shown here. It is hoped that this study will be repeated in the future as a way of continuously tracking CPUT research output. The framework presented in this article can be used by other intuitions. Such studies could create a pool of research on publication trends which will be comparable and could be used to benchmark growth.

## Acknowledgements

The authors would like to acknowledge colleagues at CPUT Libraries and CPUT's Research, Technology, Innovation and Partnerships Unit for assistance in gathering various data and information required for writing the article.

## Author Contributions

E.C. was the project leader and provided the framework, the content of the introduction and parts of the literature. L.S. provided the initial analyses and interpretations. Both worked on the editing.

## References

1. Cape Peninsula University of Technology. Research technology innovation and partnerships: 10-year blueprint [document on the Internet]. c2012 [cited 2015 Feb 16]. Available from: [http://www.cput.ac.za/storage/research/RTI\\_Blueprint\\_2013\\_WEB.pdf](http://www.cput.ac.za/storage/research/RTI_Blueprint_2013_WEB.pdf)
2. Baker G. Scholarly publishing in South Africa: Facing reality. *S Afr J Sci.* 2008;104:411–412. Available from: [http://www.scielo.org.za/scielo.php?pid=S0038-23532008000600009&script=sci\\_arttext&tng=es](http://www.scielo.org.za/scielo.php?pid=S0038-23532008000600009&script=sci_arttext&tng=es)
3. Schemm Y. Africa doubles research output over past decade, moves towards a knowledge-based economy. *Res Trends.* 2013;35. Available from: <http://www.researchtrends.com/issue-35-december-2013/africa-doubles-research-output/>
4. Li EY, Liao CH, Yen HR. Co-authorship networks and research impact: A social capital perspective. *Res Policy.* 2013;42(9):1515–1530. <http://dx.doi.org/10.1016/j.respol.2013.06.012>
5. Holmgren M, Schnitzer S. Science on the rise in developing countries. *PLoS Biol.* 2004;2(1):E1. <http://dx.doi.org/10.1371/journal.pbio.0020001>
6. Beattie V, Goodacre A. Publishing patterns within the UK accounting and finance academic community. *Br Account Rev.* 2004;36(1):7–44. <http://dx.doi.org/10.1016/j.bar.2003.08.003>
7. Pouris A. Effects of funding policies on research publications in South Africa. *S Afr J Sci.* 1991;87(3/4):78–81.
8. Ashoor MS, Chaudhry AS. Publication patterns of scientists working in Saudi Arabia. *Int Inf Libr Rev.* 1993;25(1):61–71. <http://dx.doi.org/10.1006/iilr.1993.1005>
9. Tijssen RJ. Africa's contribution to the worldwide research literature: New analytical perspectives, trends, and performance indicators. *Scientometrics.* 2007;71(2):303–327. <http://dx.doi.org/10.1007/s11192-007-1658-3>
10. Russell JM. Publishing patterns of Mexican scientists: Differences between national and international papers. *Scientometrics.* 1998;41(1/2):113–124. <http://dx.doi.org/10.1007/BF02457972>
11. Kotecha P, Walwyn D, Pinto C. Deepening research capacity and collaboration across universities in SADC: A southern African universities regional research and development fund [document on the Internet]. c2011 [cited 2015 Feb 16]. Available from: [http://www.sarua.org/files/publications/RDFund/SARU\\_RD%20Fund\\_May%202011%20Ver%201.pdf](http://www.sarua.org/files/publications/RDFund/SARU_RD%20Fund_May%202011%20Ver%201.pdf)
12. Mali F, Kronegger L, Ferligoj A. Co-authorship trends and collaboration patterns in the Slovenian sociological community. *Corvinus J Sociol Soc Policy.* 2010;1(2):29–50. Available from: <http://cjssp.uni-corvinus.hu/index.php/cjssp/article/view/24>
13. Cantner U, Rake B. International research networks in pharmaceuticals: Structure and dynamics. *Res Policy.* 2014;43(2):333–348. <http://dx.doi.org/10.1016/j.respol.2013.10.016>
14. Ynalvez MA, Shrum WM. Professional networks, scientific collaboration, and publication productivity in resource-constrained research institutions in a developing country. *Res Policy.* 2011;40(2):204–216. <http://dx.doi.org/10.1016/j.respol.2010.10.004>
15. Katz JS, Martin BR. What is research collaboration? *Res Policy.* 1997;26(1):1–18. [http://dx.doi.org/10.1016/S0048-7333\(96\)00917-1](http://dx.doi.org/10.1016/S0048-7333(96)00917-1)
16. Glänzel W, Schubert A. Analysing scientific networks through co-authorship [chapter on the Internet]. In: Moed HF, Glänzel W, Schmoch U, editors. *Handbook of quantitative science and technology research.* Amsterdam: Springer; 2005 p. 257–276. [http://dx.doi.org/10.1007/1-4020-2755-9\\_12](http://dx.doi.org/10.1007/1-4020-2755-9_12)
17. Toral SL, Bessis N, Del Rocio Martínez-Torres M. External collaboration patterns of research institutions using shared publications in the Web of Science. *Program.* 2013;47(2):170–187. <http://dx.doi.org/10.1108/00330331311313753>
18. Newman MEJ. Coauthorship networks and patterns of scientific collaboration. *Proc Natl Acad Sci USA.* 2004;101(s1):5200–5205. <http://dx.doi.org/10.1073/pnas.0307545100>
19. Bidault F, Hildebrand T. The distribution of partnership returns: Evidence from co-authorships in economics journals. *Res Policy.* 2014;43(6):1002–1013. <http://dx.doi.org/10.1016/j.respol.2014.01.008>
20. Wagner CS, Leydesdorff L. Network structure, self-organization, and the growth of international collaboration in science. *Res Policy.* 2005;34(10):1608–1618. <http://dx.doi.org/10.1016/j.respol.2005.08.002>

21. Pouris A, Ho YS. Research emphasis and collaboration in Africa. *Scientometrics*. 2014;98(3):2169–2184. <http://dx.doi.org/10.1007/s11192-013-1156-8>
22. 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. <http://dx.doi.org/10.1007/s11192-010-0330-5>
23. Boshoff N. South–South research collaboration of countries in the Southern African Development Community (SADC). *Scientometrics*. 2009;84(2):481–503. <http://dx.doi.org/10.1007/s11192-009-0120-0>
24. Sooryamoorthy R. Scientific collaboration in South Africa. *S Afr J Sci*. 2013;109(5/6), Art. #a0016, 5 pages. <http://dx.doi.org/10.1590/sajs.2013/a0016>
25. Sooryamoorthy R. Collaboration and publication: How collaborative are scientists in South Africa? *Scientometrics*. 2009;80(2):419–439. <http://dx.doi.org/10.1007/s11192-008-2074-z>
26. Barnard H, Cowan R, Müller M. Global excellence at the expense of local diffusion, or a bridge between two worlds? *Research in science and technology in the developing world. Res Policy*. 2012;41(4):756–769. <http://dx.doi.org/10.1016/j.respol.2011.12.002>
27. Dell S. International collaboration in African research: Who wins? [article on the Internet]. *Univ World News*. 2014;354. Available from: <http://www.universityworldnews.com/article.php?story=2014020317280463>
28. Sooryamoorthy R. Science and scientific collaboration in South Africa: Apartheid and after. *Scientometrics*. 2009;84(2):373–390. <http://dx.doi.org/10.1007/s11192-009-0106-y>
29. Sooryamoorthy R. The visibility of engineering research in South Africa. *S Afr J Ind Eng*. 2010;21(2):1–12. Available from: [http://www.scielo.org.za/scielo.php?pid=S2224-78902010000200002&script=sci\\_arttext](http://www.scielo.org.za/scielo.php?pid=S2224-78902010000200002&script=sci_arttext)
30. Leiden University: Centre for Science and Technological Studies. VOSViewer [programme on the Internet]. c2013 [cited 2014 Apr 30]. Available from: <http://www.vosviewer.com/>
31. 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. <http://dx.doi.org/10.4102/sajs.v107i1/2.406>
32. Pouris A. Scientometric research in South Africa and successful policy instruments. *Scientometrics*. 2012;91(2):317–325. <http://dx.doi.org/10.1007/s11192-011-0581-9>
33. MacRoberts M, MacRoberts B. Problems of citation analysis. *Scientometrics*. 1996;36(3):435–444. Available from: <http://www.akademai.com/doi/abs/10.1007/BF02129604>



# Physicochemical characteristics of Bambara groundnut dietary fibres extracted using wet milling

## AUTHORS:

Yvonne Maphosa<sup>1</sup>  
Victoria. A. Jideani<sup>1</sup>

## AFFILIATION:

<sup>1</sup>Department of Food Technology,  
Cape Peninsula University  
of Technology, Cape Town,  
South Africa

## CORRESPONDENCE TO:

Yvonne Maphosa

## EMAIL:

yvonmaphosa@gmail.com

## POSTAL ADDRESS:

Department of Food Technology,  
Cape Peninsula University of  
Technology, PO Box 1906,  
Bellville 7535, South Africa

## DATES:

Received: 30 Mar. 2015

Revised: 17 May 2015

Accepted: 06 July 2015

## KEYWORDS:

swelling capacity; bulk density;  
water holding capacity;  
soluble dietary fibre; insoluble  
dietary fibre

## HOW TO CITE:

Maphosa Y, Jideani  
VA. Physicochemical  
characteristics of Bambara  
groundnut dietary fibres  
extracted using wet milling.  
S Afr J Sci. 2016;112(1/2),  
Art. #2015-0126, 8 pages.  
[http://dx.doi.org/10.17159/  
sajs.2016/20150126](http://dx.doi.org/10.17159/sajs.2016/20150126)

The objectives of this study were to extract soluble and insoluble dietary fibres from four Bambara groundnut (BGN) varieties (black-eye, brown-eye, brown and red) using the wet milling method and evaluate their physicochemical properties. The swelling capacities of brown-eye (6.5 g/mL) and black-eye (6.2 g/mL) fibres were higher ( $p \leq 0.05$ ) than those of red (6.0 g/mL) and brown (5.5 g/mL) fibres while the water holding capacities of black-eye and brown-eye fibres (2.84 g and 2.83 g water/g sample) were higher ( $p \leq 0.05$ ) than those of brown and red fibres. The bulk densities of insoluble dietary fibres (IDFs) and soluble dietary fibres (SDFs) ranged between 0.57 g/mL (red) to 0.67 g/mL (brown-eye) and 0.46 g/mL (brown-eye) to 0.57 g/mL (black-eye), respectively. The oil binding capacities (OBCs) of SDFs ranged between 2.78 g oil/g sample (brown) and 4.03 g oil/g sample (brown-eye) while the OBC of all IDFs did not differ ( $p > 0.05$ ), ranging between 1.52 g oil/g sample (brown) and 1.40 g oil/g sample (brown-eye and black-eye). Black-eye and brown-eye dietary fibres had higher phenolic and total sugar content. The findings of this study indicate the potential of BGN fibres in food systems as fat replacers, emulsion stabilisers, water binders, bulking agents, thickeners and nutritional additives.

## Introduction

Bambara groundnut (BGN) is an underutilised crop predominantly grown in African countries.<sup>1,2</sup> Legume seeds such as BGN are good sources of dietary fibre<sup>3</sup> and BGN fibre has potential for both food and non-food applications.<sup>4</sup> An increase in consumer awareness of the health benefits of dietary fibre (DF) has led to the investigation of alternative sources by a number of researchers.<sup>5-7</sup> These health benefits include reduced risk of diseases of lifestyle, such as obesity, diabetes, coronary heart disease, some cancers and haemorrhoids.<sup>7-9</sup>

Legumes that have been researched for DF extraction include cowpeas, lentils and chickpeas.<sup>5,10</sup> The basis of DF extraction methods is similar, however, the approach differs depending on the desired end product, source of fibre and availability of equipment. All DF extraction methods involve fractionation as this allows for the separation of constituents to obtain the desired concentrates and isolates.<sup>11</sup> Some methods of extracting DF include microbiological retting, chemical, enzymatic, dry processing and wet processing.<sup>12</sup>

The modified wet milling method as reported by Dalgetty and Baik<sup>6</sup> is more efficient than the conventional wet methods that rely solely on the differences in swelling capacity to separate starch and fibres, as it makes use of the enzyme  $\alpha$ -amylase to digest any remaining starch, thus purifying the fibre concentrate. Extracted DF from BGN using the wet milling method is not documented. Furthermore, the properties and applications of BGN DFs are not largely documented. An understanding of the physicochemical properties of BGN DF will highlight the behaviour in different food and non-food systems including in the human gastro-intestinal tract.<sup>13,14</sup> Diedericks<sup>15</sup> applied an enzymatic-gravimetric method of extracting DF from BGN. The method proved to be very costly and time consuming costing approximately ZAR26 388.57/kg DF. The wet milling method is cheaper and easy to handle.<sup>6</sup> Therefore, the objectives of this study were to extract soluble and insoluble fibres from whole seeds of BGN varieties using the wet milling method as an alternative to the enzymatic-gravimetric method and evaluate their physicochemical properties.

## Materials and Methods

### Materials

BGN seeds were purchased from Triotrade in Johannesburg, South Africa, and sorted into four varieties according to the 'eye' colour, namely, the black-eye, brown-eye, brown and red varieties. Chemicals used in this study were of analytical grade (Sigma-Aldrich, Johannesburg, South Africa). Equipment used was obtained from the Departments of Food Technology and Oxidative Stress of the Cape Peninsula University of Technology.

### Milling of Bambara groundnut seeds

BGN seeds were washed and then dried at 50 °C for 48 h (Cabinet drier, Model: 1069616, Geiger & Klotzbucher, Cape Town, South Africa). The seeds were then milled using a hammer mill (Bauermeister, Bauermeister Inc., Vernon Hills, IL, USA) with a sieve size of 250  $\mu$ m.

### Wet fractionation of BGN flour into individual constituents

The method of Dalgetty and Baik<sup>6</sup> was adopted in this study. BGN flour (200 g) was mixed with 500 mL distilled water and blended for 3 min at the highest setting. The slurry was centrifuged (15 min, 25 °C, 1500 x g). The residue was used for the isolation of insoluble dietary fibre (IDF) and the supernatant was used in the isolation of soluble dietary fibre (SDF).

### Isolation of BGN insoluble dietary fibre

The residue (26 g) was wet screened in 2 L of water through a 53  $\mu\text{m}$  sieve. The supernatant was collected as starch concentrate. To purify IDF, the collected sediment was digested using 13 units/mg 10 MU heat-stable  $\alpha$ -amylase in 400 mL of water at pH 6 for 30 min, in a shaking water bath (100 °C). After digestion, the tubes were left to cool down to room temperature and centrifuged (10 min, 25 °C, 1500 x g). The residue was collected and dried at 50 °C (Cabinet drier, Model: 1069616) for 48 h and then vacuum dried in an air oven at 100 °C for 2.5 h.

### Isolation of soluble dietary fibre

Soluble dietary fibre was isolated from the supernatant collected after wet fractionation. Firstly, proteins were precipitated by adjusting the pH of the soluble fraction from pH 3 to pH 9 using 1 N NaOH and 1 N HCl. Following precipitation, the soluble fraction was centrifuged (10 min, 25 °C, 1500 x g). The sediment was collected as protein concentrate. The supernatant was subjected to a tangential flow filtration system (Spectrum Laboratories Inc., Rancho Dominguez, CA, USA) and each fibre solution was washed with four diafiltration volumes to remove any contaminants. Waste was removed through a hollow fibre filtration outlet with a molecular weight cut-off of 10 kD.

## Assessment of the physicochemical properties of BGN fibres

### Hydration properties

The swelling capacities of BGN IDFs were determined using the method of Wang and Toews.<sup>16</sup> Dry, purified IDF (0.2 g) was hydrated with 10 mL of distilled water in a graduated cylinder and left to stand at 24 °C for 18 h. The swelling capacity of the fibres was then calculated as the bed volume occupied by fibres per gram of dry sample.

The method described by Dalgetty and Baik<sup>5</sup> was applied with modifications in the determination of the water holding capacity (WHC) of BGN IDFs. In a 50 mL centrifuge tube, 1 g of fibre and 30 mL of distilled water were added and the tubes were held for 18 h at 24 °C to allow sufficient hydration of the fibre. The tubes were then centrifuged (3000 x g, 20 min, 23 °C). The supernatant was decanted and the tubes carefully inverted for 10 min to drain any remaining free water. The weight of the residue was then recorded and the difference between the original volume of water and the volume of the supernatant was calculated to determine the WHC. The WHC was expressed as mL/g.

### Density of BGN fibres

The method of Parrott and Thrall<sup>17</sup> was followed in determining both bulk and direct densities. Bulk density was determined by adding 2 g of each BGN fibre into a graduated syringe and manually applying sufficient pressure while gently tapping the syringe on a bench until the contents were packed tightly. Direct density was determined by adding fibre to the 5 mL mark in a 10 mL graduated cylinder. Care was taken to avoid shaking the cylinder so as to avoid settling of the fibre.<sup>5</sup> The dietary fibre was then emptied and weighed. Bulk and direct densities were expressed in g/mL.

### Oil binding capacity of BGN fibres

The method described by Dalgetty and Baik<sup>5</sup> was applied to determine the oil binding capacity (OBC) of the BGN fibres with modifications. Fibre (1 g) was mixed with 5 g of canola oil in a 50 mL centrifuge tube. The mixture was vortexed for 30 sec at 5 min intervals for 30 min. The mixture was then centrifuged (1600 x g, 25 min, 23 °C). After centrifugation, the supernatant (free oil) was decanted and weighed. OBC was expressed as grams of oil retained/grams of fibre.

### Colour measurements of BGN fibres

Colour attributes of BGN dietary fibres were determined using a spectrophotometer (Model CM-5, Konica Minolta Sensing, Osaka, Japan) set at standard observer 10° and D65. The spectrophotometer was calibrated with black and white plates followed by zero calibration. BGN fibres (3 g of IDF and 0.6–0.8 g SDF) were placed in a glass sample

holder (diameter 30 mm). Lightness ( $L^*$ ), redness/greenness ( $a^*$ ) and yellowness/blueness ( $b^*$ ), hue and chromacity were assessed through  $L^*C^*h^*$  and CIE- $L^*a^*b^*$  colour space systems. Each variety was analysed in triplicate with each individual sample giving three readings. Colour differences amongst the fibre samples were calculated using the colour difference equation:

$$\Delta E = \sqrt{(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2} \quad \text{Equation 1}$$

where  $L^*$  is lightness,  $a^*$  is redness/greenness and  $b^*$  is yellowness/blueness

### Assessment of polyphenolic compounds in BGN fibres

The method of Diedericks<sup>15</sup> was adopted in the assessment of polyphenolic compounds. Condensed tannins were determined in IDFs while hydrolysable polyphenols (HPPs) were determined in both SDFs and IDFs. For determination of HPPs in BGN IDFs, samples (250 mg) were mixed with 10 mL of methanol and 1 mL of  $\text{H}_2\text{SO}_4$  in 14 mL centrifuge tubes. The samples were incubated at 80 °C for 20 h. The samples were then centrifuged (4000 x g, 5 min, 21 °C) and the residues were analysed using the Folin–Ciocalteu assay by mixing 25  $\mu\text{L}$  of sample with 125  $\mu\text{L}$  of 0.2 M Folin–Ciocalteu and 100  $\mu\text{L}$  of 7.5%  $\text{Na}_2\text{CO}_3$  solution. The mixtures were left to stand for 2 h then the absorbance was measured using a spectrophotometer at 750 nm using a gallic acid standard calibration curve. The results were expressed as mg/g gallic acid equivalents (GAE). For determination of HPPs, BGN SDF samples (250 mg) were dissolved in 10 mL distilled water, centrifuged (4000 x g, 5 min, 21 °C) and the supernatant was subjected to the Folin–Ciocalteu assay. Tannins were determined by treating IDF samples (250 mg) with a 1:1 mixture of 5 mL/L HCl-Butanol. The mixture was incubated at 100 °C for 1 h then centrifuged (4000 x g, 5 min, 21 °C). Tannins were calculated from the anthocyanidin solutions absorbance at a wavelength of 555 nm using a standard curve of 0.0072 ppm and an absorbance of +0.0072.

### Assessment of neutral sugars and uronic acids in BGN fibres

BGN fibres were subjected to acid hydrolysis prior to analysis of neutral sugars and uronic acids. SDFs were hydrolysed with 1 M  $\text{H}_2\text{SO}_4$  at 100 °C for 90 min and IDFs were first hydrolysed with 12 M  $\text{H}_2\text{SO}_4$  at 30 °C for 90 min and then with 1 M  $\text{H}_2\text{SO}_4$  at 100 °C for 90 min to yield monomers. After hydrolysis, samples were centrifuged (3000 x g, 15 min, 21 °C). IDF residues were washed twice with 2 mL distilled water and SDFs were filtered to remove any suspensions. Uronic acids and neutral sugars were then analysed in the IDF and SDF supernatants by spectrophotometry (340 nm) using K-Arga, K-Fucose, K-Mangl, K-Rhan, K-Uronic and K-Xylose assay kits (Megazyme International, Wicklow, Ireland).

### Data analysis

For statistical analysis, IBM Statistical Package for the Social Science (IBM SPSS, version 22) was used. The results were subjected to Multivariate Analysis of Variance (MANOVA) to determine mean differences between treatments. Duncan's multiple range test was conducted to separate mean differences where differences existed.

## Results and Discussion

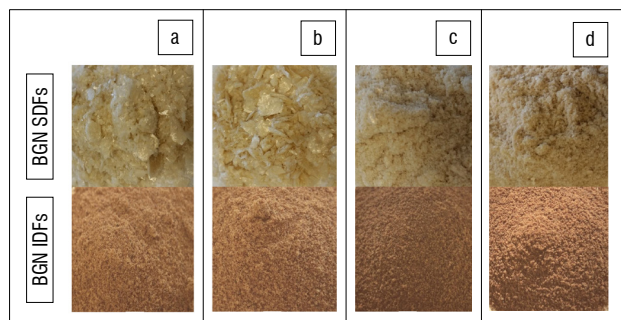
### Yield of BGN fibres

Soluble and insoluble dietary fibres were successfully isolated from four varieties of BGN using the modified wet milling method (Figure 1) and the yield of each dietary fibre is given in Table 1. Scanning electron micrographs of BGN fibres are shown in Figure 2. The yield of SDFs was in the range 15.4% (red) to 17.1% (brown-eye) and that of IDFs was in the range 12.0% (brown-eye) to 15.6% (red). There was no significant difference ( $p > 0.05$ ) in the yield of SDFs as well as among the IDFs.

Using the enzymatic-gravimetric method to extract DFs from legumes, a lower yield of both IDFs and SDFs has been reported. The lower yield of DFs obtained using the enzymatic-gravimetric-method may be attributed



to the fact that chemicals used in this method result in the loss of some IDFs and most SDFs.<sup>18</sup> The yield of BGN DFs in this study was considered high as several researchers have reported BGN DF content in the range 5.2% to 6.4%.<sup>19-21</sup> The variations in yield among researchers may be attributed to differences in BGN varieties, climatic conditions, type of soil grown on, processing and determination methods used. The SDF content of most edible legumes such as pea, broad pea and soybean cotyledons range between 3.3% and 13.8%.<sup>10,22</sup> The yield of SDFs in this study was higher than the reported range. This increase in yield could be an indication that BGN has a higher SDF content than previously studied legumes. In addition, the use of different extraction methods could be responsible for the differences in yield; the wet milling method could have favoured the extraction of soluble fractions of DF more than other methods.



**Figure 1:** Soluble and insoluble dietary fibres (SDFs and IDFs, respectively) isolated from four varieties of Bambara groundnut (BGN): (a) black-eye, (b) brown-eye, (c) brown and (d) red.

### Hydration properties of BGN dietary fibres

#### Swelling capacity of BGN fibres

The swelling capacity of IDFs ranged from 6.50 mL/g (brown-eye) to 5.50 mL/g (brown) as shown in Table 1. There were no significant ( $p > 0.05$ ) differences between brown and red IDFs and also among black-eye, brown-eye and red IDFs in terms of swelling capacity. Brown-eye and black-eye IDFs had higher ( $p \leq 0.05$ ) swelling capacities than brown and red IDFs. Swelling capacities in the range 4.28–5.51 mL/g were reported for chickpea and pea IDFs.<sup>5,23</sup> These values are comparable to those obtained in this study. A swelling capacity of 5.51 mL/g was reported for mung bean hulls sieved through a 50  $\mu\text{m}$  mesh which is similar to the sieve size used in this study.<sup>24</sup> The similarities in the results of the two studies suggest that particle size plays a major role in determining the

swelling capacity of fibres.<sup>24</sup> Increasing swelling capacity with decreasing particle size has been reported.<sup>24</sup> Thus it can be deduced that particle size has an inverse relationship with swelling capacity.

The swelling capacities of BGN fibres are comparable to that of cellulose (6.2 mL/g), a dietary fibre constituent that is widely used in food products as a bulking agent, stabiliser, thickener and anti-caking agent owing to its hydration properties.<sup>25</sup> Reduced cooking losses, decreased firmness, decreased adhesiveness and reduced stickiness in pea and inulin fibre enriched pastas have been reported.<sup>26</sup> The researchers reported that the increase in swelling capacities of the fibres imparted these desirable characteristics in the pastas.

This study suggested that BGN fibres would make suitable substitutes for cellulose, inulin and pea fibres in food systems such as pasta because of their swelling capacities. Physiologically, the swelling capacity of fibres is important in the control of blood glucose levels and also contributes to proper gut function.<sup>7,9</sup>

#### Water holding capacity of BGN fibres

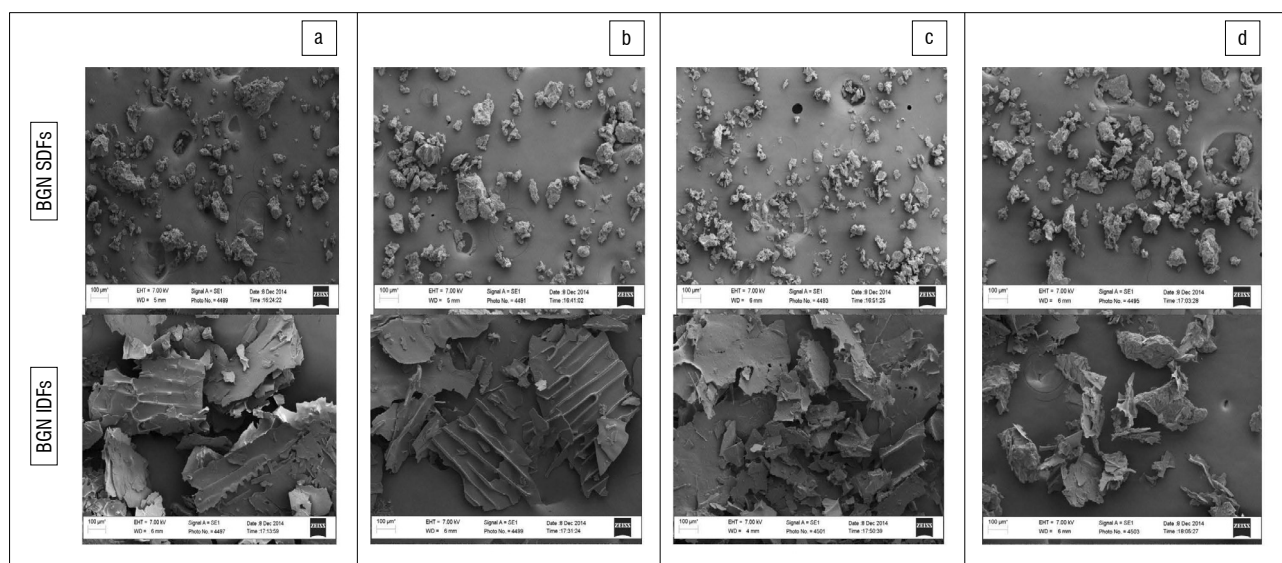
The water holding capacity (WHC) of BGN fibres ranged from 2.41 g water/g sample (red) to 2.84 g water/g sample (black-eye) as shown in Table 1. The WHCs of black-eye and brown-eye IDF were significantly ( $p \leq 0.05$ ) higher than the WHC of brown IDF and brown IDF WHC in turn was higher ( $p \leq 0.05$ ) than red IDF WHC.

The WHC of passion fruit seed IDF was reported as 2.37 g water/g sample.<sup>27</sup> Passion fruit seed fibre has been described as a functional ingredient that improves the health and functioning of the gut attributing these characteristics to its WHC.<sup>28,29</sup> As the WHC of BGN IDF is comparable to that of passion fruit seed IDF, it can be deduced that BGN IDF can play a similar physiological role.

Several researchers reported the WHC of various legumes in the range 3.13 g water/g sample to 13.4 g water/g sample.<sup>5,13,23,24</sup> The higher values obtained by these researchers could be because of different particle sizes, as well as differences in the chemical nature, composition and processing history of the fibres.<sup>30</sup> The differences can also be attributed to the difference in legume species. The WHC of BGN fibres could find use in the meat, dairy and bakery industries.<sup>31</sup>

#### Densities of BGN fibres

The bulk and direct densities of BGN SDFs and IDFs were evaluated and the findings are given in Table 1. SDFs had bulk densities in the range 0.46 g/mL (black-eye and brown-eye) to 0.57 g/mL (brown) and direct



**Figure 2:** Scanning electron micrographs of soluble and insoluble dietary fibres (SDFs and IDFs, respectively) isolated from four varieties of Bambara groundnut (BGN). (a) black-eye, (b) brown-eye, (c) brown, (d) red.

Table 1: Physicochemical properties of Bambara groundnut dietary fibres

	Yield (%)	Swelling capacity (mL/g)	Water holding capacity (g/g)	Direct density (g/mL)	Bulk density (g/mL)	Oil binding capacity (g/g)	Hydrolysable polyphenols (mg/g)	Tannins (mg/g)
<b>Soluble dietary fibre</b>								
Black-eye	15.4 ± 1.7 <sup>a</sup>	-	-	0.05 ± 0.0 <sup>a</sup>	0.46 ± 0.0 <sup>a</sup>	3.84 ± 0.0 <sup>a</sup>	19.87 ± 0.5 <sup>a</sup>	-
Brown-eye	17.1 ± 3.3 <sup>a</sup>	-	-	0.06 ± 0.0 <sup>a</sup>	0.46 ± 0.0 <sup>a</sup>	4.03 ± 0.1 <sup>b</sup>	20.86 ± 0.0 <sup>b</sup>	-
Brown	15.4 ± 0.9 <sup>a</sup>	-	-	0.11 ± 0.0 <sup>a</sup>	0.57 ± 0.0 <sup>a</sup>	3.72 ± 0.1 <sup>a</sup>	6.89 ± 0.2 <sup>c</sup>	-
Red	17.1 ± 0.6 <sup>a</sup>	-	-	0.07 ± 0.0 <sup>a</sup>	0.50 ± 0.0 <sup>a</sup>	2.78 ± 0.1 <sup>c</sup>	8.26 ± 0.3 <sup>d</sup>	-
<b>Insoluble dietary fibre</b>								
Black-eye	13.1 ± 2.5 <sup>a</sup>	6.17 ± 0.3 <sup>a</sup>	2.84 ± 0.1 <sup>a</sup>	0.53 ± 0.0 <sup>a</sup>	0.57 ± 0.0 <sup>ab</sup>	1.40 ± 0.7 <sup>a</sup>	10.96 ± 0.2 <sup>a</sup>	2.10 ± 0.0 <sup>a</sup>
Brown-eye	12.0 ± 1.4 <sup>a</sup>	6.50 ± 0.0 <sup>a</sup>	2.83 ± 0.1 <sup>a</sup>	0.50 ± 0.0 <sup>ab</sup>	0.67 ± 0.0 <sup>c</sup>	1.40 ± 0.2 <sup>a</sup>	11.44 ± 0.3 <sup>a</sup>	2.07 ± 0.1 <sup>a</sup>
Brown	15.6 ± 0.9 <sup>a</sup>	5.50 ± 0.0 <sup>b</sup>	2.60 ± 0.0 <sup>b</sup>	0.45 ± 0.0 <sup>c</sup>	0.57 ± 0.0 <sup>a</sup>	1.46 ± 0.6 <sup>a</sup>	14.43 ± 0.6 <sup>a</sup>	1.19 ± 0.6 <sup>b</sup>
Red	13.7 ± 2.7 <sup>a</sup>	6.00 ± 0.5 <sup>ab</sup>	2.41 ± 0.1 <sup>c</sup>	0.49 ± 0.0 <sup>b</sup>	0.58 ± 0.0 <sup>b</sup>	1.52 ± 0.0 <sup>a</sup>	13.27 ± 1.6 <sup>a</sup>	1.36 ± 1.6 <sup>c</sup>

Values are mean ± standard deviation. Means within a column followed by the same superscript are not significantly ( $p > 0.05$ ) different.

densities in the range 0.05 g/mL (black-eye) to 0.11 g/mL (brown). Red SDF was significantly ( $p \leq 0.05$ ) higher than both brown-eye and black-eye SDFs but lower ( $p \leq 0.05$ ) than brown SDF in terms of bulk density while all SDFs differed significantly ( $p \leq 0.05$ ) in direct density. IDFs had bulk densities in the range 0.57 g/mL (red and black-eye) to 0.67 g/mL (brown-eye) and direct densities in the range 0.45 g/mL (brown) to 0.53 g/mL (black-eye). The IDFs of black-eye and brown-eye as well as black-eye and red varieties did not differ significantly ( $p > 0.05$ ) in their direct densities. Brown IDF had a significantly ( $p \leq 0.05$ ) lower direct density than the other three fibres.

Direct density is measured without compressing the fibres while bulk density is measured after compressing the fibres. Consequently, bulk density measurements yielded higher values than direct densities. Dalgetty and Baik<sup>6</sup> concluded that SDFs have higher densities than IDFs. The results obtained in the current study disagree with Dalgetty and Baik's<sup>6</sup> conclusion as IDFs had higher bulk and direct densities than SDFs. The bulk densities of BGN IDFs were comparable to those of IDFs from passion fruit seeds (0.68 g/mL), soybean (0.43 g/mL), peas (0.54–0.56 g/mL), pigeon pea (0.47 g/mL) and chickpea (0.65 g/mL).<sup>27,32</sup> These fibres are commercially available hence this is one of the indications that BGN dietary fibres have the potential to successfully compete commercially with other fibres. Differences in bulk densities with other leguminous fibres could be because of different structural compositions owing to different leguminous species.<sup>13</sup>

Density is of importance in packaging, with higher densities resulting in a reduced ability to compress. Therefore, the densities of BGN fibres could be an advantage as they will pack closely together, hence requiring less packaging material, resulting in cost saving.<sup>15</sup>

#### Oil binding capacity of BGN fibres

The oil binding capacities (OBCs) of SDFs ranged from 2.78 g oil/g sample (brown) to 4.03 g oil/g sample (brown-eye) (Table 1). Black-eye and red SDFs did not differ significantly in OBC, brown-eye SDF was significantly higher than the other three SDFs and brown SDF was significantly lower than all three SDFs. Among the IDFs, brown IDF had the highest OBC of 1.52 g oil/g sample, while brown-eye and black-eye IDFs both had the lowest OBC of 1.40 g oil/g sample. All four IDFs did not differ ( $p > 0.05$ ) in terms of OBC. SDFs showed higher OBCs than IDFs.

Higher IDF OBC values were reported for pea (6.93 g oil/g sample), chickpea (4.25 g oil/g sample) and lentil fibres (4.01 g oil/g sample).<sup>5</sup> These differences can be attributed to structural differences and compositional variation of the fibres. OBC of 1.49 g oil/g sample to

1.83 g oil/g sample were reported for mung bean hulls which compares fairly with the OBC of BGN IDFs.<sup>24</sup>

A total of 11 commercial fibres were studied and low OBC values were reported with the highest being 0.02 g oil/g sample from bamboo.<sup>25</sup> The low OBC values indicated that BGN fibres can compete commercially with other fibres in stabilising high fat foods and emulsions.<sup>23</sup> Lower OBC values for SDFs derived from pea (1.15 g oil/g sample), chickpea (1.14 g oil/g sample) and lentil fibres (0.89 g oil/g sample) have been reported.<sup>5</sup>

This study indicated that the use of BGN fibres would be economical as less BGN fibre would be used to render the desirable properties compared to other leguminous fibres. The ability of fibres to bind oil can be harnessed by the food industry to reduce fat losses upon cooking and in stabilising emulsions.<sup>23,33</sup> Physiologically, the OBCs of BGN fibres would allow them to play a role in bile acid absorption and consequently cholesterol reduction.<sup>30</sup> OBC would also be significant in reducing fat absorption by the body.

#### Colour characteristics of BGN fibres

The colour attributes of BGN measured were lightness ( $L^*$ ), redness ( $+a^*$ ) / greenness ( $-a^*$ ), blueness ( $-b^*$ ) / yellowness ( $+b^*$ ), hue and chroma (Table 2). SDFs were lighter than IDFs and all the BGN fibres had  $+a^*$  and  $+b^*$  values indicating that they were more associated with redness and yellowness. The redness and yellowness of these fibres suggests their antioxidant properties.<sup>34,35</sup> BGN fibres are high in polyphenolic compounds (Table 2). Chroma describes the vividness or dullness of a colour and hue is how the colour of an object is perceived.<sup>34</sup> The hue angle of BGN DFs indicated a yellowish-red colour associated with these fibres.

Lightness of dietary fibres is of importance in food products as it determines the extent to which the original colour of the food is affected.<sup>25,30</sup> The varying colours of the BGN dietary fibres are advantageous as the manufacturer will have a choice of a fibre variety that best suits the colour of their product. IDFs had darker colours hence could find use in products such as meat emulsions and brown bread, while the lighter coloured SDFs could be used in food products such as white bread and beverage emulsions.

A colour difference ( $\Delta E$ ) of 1 is the threshold at which a trained observer would notice the difference between two colours, a  $\Delta E$  between 4 and 8 is deemed acceptable and above 8 is deemed unacceptable and likely to be rejected by consumers.<sup>36</sup> Table 3 gives the colour difference between BGN fibres. All the SDFs showed acceptable differences with  $\Delta E$  ranging between 0.81 and 3.08.

**Table 2:** Colour attributes of Bambara groundnut dietary fibres

	Lightness	Redness/greenness	Yellowness/blueness	Chroma	Hue (°)
<b>Insoluble dietary fibre</b>					
Black-eye	36.6 ± 0.4 <sup>a</sup>	9.9 ± 0.1 <sup>a</sup>	17.6 ± 0.3 <sup>a</sup>	20.2 ± 0.3 <sup>a</sup>	60.6 ± 0.3 <sup>a</sup>
Brown-eye	37.9 ± 0.5 <sup>b</sup>	10.1 ± 0.1 <sup>a</sup>	18.5 ± 0.2 <sup>b</sup>	21.1 ± 0.2 <sup>a</sup>	61.5 ± 0.2 <sup>b</sup>
Brown	24.3 ± 0.1 <sup>c</sup>	6.0 ± 0.1 <sup>b</sup>	7.9 ± 0.1 <sup>c</sup>	10.0 ± 0.1 <sup>b</sup>	52.8 ± 0.4 <sup>c</sup>
Red	30.8 ± 0.2 <sup>d</sup>	7.9 ± 0.1 <sup>c</sup>	10.5 ± 0.2 <sup>d</sup>	12.7 ± 0.9 <sup>c</sup>	52.8 ± 0.1 <sup>c</sup>
<b>Soluble dietary fibre</b>					
Black-eye	73.0 ± 0.2 <sup>ab</sup>	1.7 ± 0.0 <sup>a</sup>	13.8 ± 0.1 <sup>a</sup>	13.9 ± 0.1 <sup>a</sup>	83.1 ± 0.1 <sup>a</sup>
Brown-eye	74.0 ± 0.5 <sup>a</sup>	2.4 ± 0.0 <sup>b</sup>	15.5 ± 0.4 <sup>b</sup>	15.7 ± 0.1 <sup>b</sup>	81.2 ± 0.5 <sup>ab</sup>
Brown	71.7 ± 1.1 <sup>bc</sup>	2.3 ± 0.3 <sup>b</sup>	15.5 ± 0.3 <sup>b</sup>	15.9 ± 0.8 <sup>b</sup>	81.1 ± 1.8 <sup>ab</sup>
Red	71.0 ± 1.3 <sup>c</sup>	2.5 ± 0.2 <sup>b</sup>	15.6 ± 0.5 <sup>b</sup>	16.3 ± 1.0 <sup>b</sup>	79.8 ± 1.4 <sup>b</sup>

Values are mean ± standard deviation. Means within a column followed by the same superscript are not significantly ( $p > 0.05$ ) different.

Hence, they could be used interchangeably in products without a noticeable difference. IDFs, with the exception of black-eye – brown-eye and brown – red comparisons had  $\Delta E$  above 8 meaning their colour differences were very apparent and if used interchangeably, a perceivable difference would be expected.

**Table 3:** Colour differences ( $\Delta E$ ) between Bambara groundnut dietary fibres

Variety	Insoluble dietary fibre $\Delta E$	Soluble dietary fibre $\Delta E$
Black-eye – brown-eye	1.55	2.12
Black-eye – brown	16.14	2.22
Black-eye – red	18.66	2.81
Brown-eye – brown	17.67	2.31
Brown-eye – red	10.94	3.08
Brown – red	7.20	0.81

### Phenolic content of BGN fibres

The antioxidant capacity of BGN fibres, as represented by hydrolysable phenols (HPPs) and tannins, is shown in Table 1. The HPP content of SDFs ranged from 6.89 mg/g GAE (brown) to 20.86 mg/g GAE (brown-eye). All four SDFs differed significantly in their HPP content. The HPP content of IDFs ranged from 10.96 mg/g GAE (black-eye) to 14.43 mg/g GAE (brown). Black-eye and brown-eye IDFs as well as brown and red IDFs did not differ significantly in their HPP content. The tannin content of IDFs ranged from 1.12 mg/g (black-eye) to 2.1 mg/g (brown). Black-eye and brown-eye IDFs were both significantly higher ( $p \leq 0.05$ ) in tannin content than red and brown IDFs. Brown IDF was significantly lower than the other three fibres in terms of tannin content. The difference in phenolic content of the BGN fibres was in agreement with Nti<sup>35</sup> who reported that tannin content differs from one BGN variety to another.

The high polyphenolic composition of BGN reveals their potential antioxidant properties. It has been suggested that fibres can be exploited as novel antioxidants and would be of importance in protecting against superoxide radicals, hydroxyl free radicals and lipid peroxidation.<sup>23</sup> They would thus find use as ingredients in fatty foodstuffs to improve oxidative

stability, hence improving their shelf life. Antioxidants are important for human health as they prevent some degenerative diseases like cancers and decrease the oxidation of low density lipoproteins, thereby avoiding arteriosclerosis and related coronary heart diseases.<sup>37-39</sup> Antioxidants carry out these functions by reacting with free radicals forming stable or non-reactive radicals.<sup>39</sup> BGN fibres can be a useful source of natural antioxidants as alternatives to artificial antioxidants; artificial antioxidants have been shown to be carcinogenic and teratogenic.<sup>39</sup> The low tannin content observed in BGN fibres could be advantageous as tannins have been associated with anti-nutritional properties because of their ability to form complexes with some nutrients, including divalent minerals and proteins, rendering them bio-unavailable.<sup>15,40</sup>

### Neutral sugars and uronic acids in BGN fibres

Seven neutral sugars were analysed for in BGN fibres and the results are given in Table 4. Arabinose and galactose coeluted in this study and therefore are presented as arabinose/galactose in Table 4. The percentage of arabinose/galactose in SDFs was in the range 9.4% (brown) to 19.6% (black-eye). The percentage of fructose in SDFs ranged from 1.3% (black-eye) to 1.7% (brown). Fucose and glucose were obtained in low amounts (below 1%) in SDFs while relatively higher percentages of xylose were obtained in the range 13.0% (brown) to 16.6% (black-eye) (Table 4). Brown-eye and black-eye SDFs did not differ significantly ( $p > 0.05$ ) in their sugar composition with the exception of arabinose/galactose. Rhamnose was absent in all SDFs. This finding is in agreement with Dalgetty and Baik<sup>5</sup> who reported the absence of rhamnose in SDFs of pea, lentil and chickpea. The researchers also reported the absence of arabinose and mannose in SDFs. In the current study however, these two sugars were present. These authors further reported higher values of xylose (32%) in chickpea SDF. These differences can be attributed to the different analytical methodologies adopted, with sugar assay having been used in the current study and HPLC having been used in the study by Dalgetty and Baik.<sup>5</sup>

The presence of these sugars in BGN SDFs suggests the possible presence of galactomannans, arabinoxylans and arabinogalactans. Galactomannans are related to locust bean and guar gums and their solubility in water increases with increasing galactose content. BGN SDFs had higher quantities of galactose compared to mannose hence the solubility of galactomannans would be elevated.<sup>41</sup> Arabinoxylans possess antioxidant capabilities and influence water balance and rheology<sup>42</sup> and arabinogalactans possess similar characteristics as gum Arabic.<sup>43-45</sup> The suggestive presence of these hydrocolloids in BGN fibres could mean that BGN fibres possess similar beneficial characteristics and thus can be classified with them.



**Table 4:** Neutral sugar and uronic acid composition of Bambara groundnut dietary fibres

SUGARS (%)								
Variety	Arabinose/ Galactose	Fructose	Fucose	Glucose	Mannose	Rhamnose	Xylose	Uronic acids
<b>Soluble dietary fibre</b>								
Black-eye	19.6 ± 0.9 <sup>a</sup>	1.3 ± 0.1 <sup>a</sup>	0.3 ± 0.1 <sup>a</sup>	1.0 ± 0.1 <sup>a</sup>	6.1 ± 0.5 <sup>a</sup>	0	16.6 ± 1.6 <sup>a</sup>	11.5 ± 1.0 <sup>a</sup>
Brown-eye	15.2 ± 1.5 <sup>b</sup>	1.5 ± 0.1 <sup>ab</sup>	0.3 ± 0.1 <sup>a</sup>	0.8 ± 0.1 <sup>ab</sup>	6.5 ± 0.2 <sup>a</sup>	0	15.6 ± 0.2 <sup>ab</sup>	10.2 ± 0.8 <sup>a</sup>
Brown	9.4 ± 0.1 <sup>c</sup>	1.7 ± 0.1 <sup>b</sup>	0.1 ± 0.0 <sup>b</sup>	0.8 ± 0.1 <sup>ab</sup>	7.4 ± 0.6 <sup>b</sup>	0	13.0 ± 1.8 <sup>c</sup>	10.6 ± 0.3 <sup>a</sup>
Red	10.0 ± 0.6 <sup>c</sup>	1.6 ± 0.1 <sup>b</sup>	0.3 ± 0.0 <sup>a</sup>	0.6 ± 0.2 <sup>b</sup>	5.0 ± 0.3 <sup>c</sup>	0	13.7 ± 0.8 <sup>bc</sup>	11.5 ± 1.1 <sup>a</sup>
<b>Insoluble dietary fibre</b>								
Black-eye	2.8 ± 0.2 <sup>a</sup>	1.7 ± 0.1 <sup>ac</sup>	0.1 ± 0.0 <sup>a</sup>	0.6 ± 0.1 <sup>a</sup>	4.7 ± 0.4 <sup>a</sup>	2.6 ± 0.1 <sup>a</sup>	11.3 ± 1.2 <sup>a</sup>	6.7 ± 0.7 <sup>a</sup>
Brown-eye	2.3 ± 0.1 <sup>b</sup>	1.7 ± 0.1 <sup>a</sup>	0.3 ± 0.1 <sup>b</sup>	0.7 ± 0.1 <sup>ab</sup>	6.6 ± 0.1 <sup>ab</sup>	1.0 ± 0.1 <sup>b</sup>	11.3 ± 1.9 <sup>a</sup>	8.6 ± 0.7 <sup>b</sup>
Brown	2.5 ± 0.2 <sup>ab</sup>	1.5 ± 0.1 <sup>b</sup>	0.4 ± 0.1 <sup>b</sup>	0.9 ± 0.1 <sup>c</sup>	5.6 ± 0.4 <sup>c</sup>	1.2 ± 0.1 <sup>c</sup>	13.8 ± 2.3 <sup>a</sup>	10.5 ± 0.6 <sup>ab</sup>
Red	2.8 ± 0.2 <sup>a</sup>	1.9 ± 0.1 <sup>c</sup>	0.2 ± 0.1 <sup>a</sup>	0.8 ± 0.0 <sup>b</sup>	5.6 ± 0.6 <sup>b</sup>	1.1 ± 0.1 <sup>c</sup>	12.4 ± 0.8 <sup>a</sup>	10.6 ± 1.0 <sup>a</sup>

Values are mean ± standard deviation. Means within a column followed by the same superscript are not significantly ( $p > 0.05$ ) different.

The uronic acid content of SDFs did not differ significantly ( $p > 0.05$ ) and ranged from 10.6% (brown) to 11.5% (red). Dalgetty and Baik<sup>5</sup> reported lower uronic acids in pea, lentil and chickpea SDFs in the range 0.2% to 1.3%. Uronic acids form salts with some wastes in the human body thereby facilitating their excretion.<sup>46</sup> Hence, their presence in BGN DFs would contribute to the body's detoxification.

In IDF, arabinose/galactose was in the range 2.3% (brown-eye) to 2.8% (red). The percentage of fructose in IDFs ranged from 1.5% (brown) to 1.9% (red). Fucose and glucose were obtained in low amounts (below 1%) while rhamnose ranged from 1.0% (brown-eye) to 2.6% (black-eye). Relatively high percentages of xylose were obtained in IDFs (Table 4). The presence of these in BGN IDFs is an indication of the presence of some polysaccharides such as cellulose (glucose) and hemicellulose (xylose, glucose, arabinose, galactose and mannose).<sup>15,47</sup>

The presence of these sugars in BGN IDFs suggests the presence of some polysaccharides such as pectic substances (rhamnose and galactose)<sup>24</sup> which in turn suggests the presence of rhamnogalacturonans.<sup>15</sup> Rhamnogalacturonans have been reported to bind heavy metals in the human body as well as lower blood cholesterol.<sup>43,44</sup> The presence of arabinose and xylose in IDFs suggest the presence of arabinoxylans. Low quantities of arabinose and galactose in IDFs suggest low quantities of arabinogalactans.

The uronic acid content of IDFs ranged from 6.7% (black-eye) to 10.6% (red). There was no significant difference among the red, brown and black-eye IDFs as well as between brown and brown-eye IDFs in uronic acid content. Dalgetty and Baik<sup>5</sup> reported lower uronic acid of pea, lentil and chickpea IDFs in the range 2.0% to 2.8% indicating the superiority in uronic acid content of BGN IDFs over other leguminous IDFs.

## Conclusions

The wet milling method was successfully applied in the extraction of BGN SDFs and IDFs yielding an appreciable amount of both fractions. Black-eye and brown-eye fibres have superior physicochemical properties compared to the brown and red fibres as evidenced by their higher swelling capacities, water-holding capacities, oil binding capacities, phenolic as

well as total sugar content. Brown-eye and black-eye IDFs were lighter in colour, yellower, redder, more saturated and had higher hues compared to the red and brown IDFs. The physicochemical properties of BGN fibres make them valuable to the food industry as thickening agents, stabilisers, health ingredients as well as cryoprotectants in frozen dairy products. BGN fibres can be considered suitable alternatives for commercial fibres such as pea, chickpea and lentil fibres as they have been shown to possess similar qualities to these fibres.

## Acknowledgements

The authors acknowledge the Cape Peninsula University of Technology (CPUT) University Research Funding and the National Research Foundation for financial assistance towards the research running costs. Bambara groundnut dietary fibre is a patent of CPUT; Dietary fibre supplement: South Africa complete patent (2014/04371).<sup>48</sup>

## Authors' contributions

The authors contributed equally to the work presented in this paper. Y.M. was responsible for the experimental work and wrote the manuscript. V.A.J. supervised the project, carried out the statistical analysis and proofread the manuscript.

## References

- Jideani VA, Diedericks CF. Nutritional, therapeutic, and prophylactic properties of *Vigna subterranea* and *Moringa oleifera*. In: Oguntibeju O. editor. Antioxidant-antidiabetic agents and human health. Croatia: InTech; 2014. p. 187. <http://dx.doi.org/10.5772/57029>
- Jideani VA, Mpotokwana SM. Modelling of water absorption of Botswana Bambara varieties using Peleg's equation. J Food Eng. 2009;92(2):182–188. <http://dx.doi.org/10.1016/j.jfoodeng.2008.10.040>
- Fasoyiro SB, Ajibade SR, Omole AJ, Adeniyani ON, Farinde EO. Proximate, mineral and anti-nutritional factors of some underutilized grain legumes in South-Western Nigeria. Nutr Food Sci. 2006;38:18–23. <http://dx.doi.org/10.1108/00346650610642151>



4. Fasoyiro S, Yudi W, Taiwo K. Processing and utilization of legumes in the tropics. In: Eissa AA, editor. Trends in vital food and control engineering. Croatia: InTech; 2012.
5. Dalgetty DD, Baik B. Isolation and characterisation of cotyledon fibers from peas, lentils, and chickpeas. *Cereal Chem.* 2003;80(3):310–315.
6. Ojimelukwe PC. Sourcing and processing of legumes. In: Onwualu PA, Obasi SC, Ukpabi, UJ, editors. Nigerian agro raw materials development. Abuja: Raw Materials Research and Development Council, 2009.
7. Daou C, Zhang H. Physico-chemical properties and antioxidant activities of Dietary Fiber derived from defatted rice bran. *Adv J Food Sci Technol.* 2011;3(5):339–347.
8. Hawkes C. Uneven dietary development: Linking the policies of globalization with the nutrition transition, obesity, and diet-related chronic diseases. *Global Health.* 2006;2:4. <http://dx.doi.org/10.1186/1744-8603-2-4>
9. Wood JA, Grusak MA. Nutritional value of chickpea. In: Yadav SS, Redden R, Chen W, Sharma B, editors. Chickpea breeding and management. Trowbridge: Cromwell Press; 2007. p.119. <http://dx.doi.org/10.1079/9781845932138.005>
10. Khan AR, Alam S, Ali S, Bibi S, Khalil IA. Dietary fiber profile of food legumes. *Sarhad J Agric.* 2007;23(3):763–766.
11. USA Dry Pea and Lentil Council. Processing methods for dry peas, lentils and chickpeas. In: Pulse processing technical manual. Moscow: USA Dry Pea and Lentil Council; 2010.
12. Bogracheva T, Cserhalmi Z, Czukor B, Fornal J, Schuster-Gajzago I, Kovacs ET, et al. Processing. In: Henley CL, editor. Carbohydrates in grain legume seeds: Improving nutritional quality and agronomic characteristics. New York: CABI Publishing; 2001. p. 89–92.
13. Tiwari U, Cummins E. Functional and physicochemical properties of legume fibers. In: Tiwari BK, Gowen, A, McKenn B, editors. Pulse foods: Processing, quality and nutraceutical applications. London: Academic Press; 2011. p. 139, 144. <http://dx.doi.org/10.1016/B978-0-12-382018-1.00005-8>
14. Urriola PE, Cervantes-Pahm SK, Stein HH. Fiber in swine nutrition. In: Chiba LI, editors. Sustainable swine nutrition. West Sussex: Wiley-Blackwell; 2013. p. 259. <http://dx.doi.org/10.1002/9781118491454.ch11>
15. Diedericks CF. Functional properties of Bambara groundnut (*Vigna Subterranea* (L.) Verdc.) non-starch polysaccharides in model and food systems [MTech dissertation]. Cape Town: Cape Peninsula University of Technology; 2014.
16. Wang N, Toews, R. Certain physicochemical and functional properties of fibre fractions from pulses. *Food Res Int.* 2011;44(8):2515–2523. <http://dx.doi.org/10.1016/j.foodres.2011.03.012>
17. Parrott ME, Thrall BE. Functional properties of various fibers: Physical properties. *J Food Sci.* 1978;43:759–763. <http://dx.doi.org/10.1111/j.1365-2621.1978.tb02412.x>
18. Gordon DT, Okuma K. Determination of total dietary fiber in selected foods containing resistant maltodextrin by enzymatic-gravimetric method and liquid chromatography: Collaborative study. *J AOAC Int.* 2002;85:435–444.
19. Baryeh EA. Physical properties of Bambara groundnuts. *J Food Eng.* 2001;47:321–326. [http://dx.doi.org/10.1016/S0260-8774\(00\)00136-9](http://dx.doi.org/10.1016/S0260-8774(00)00136-9)
20. Mkandawire CH. Review of Bambara groundnut (*Vigna subterranea* (L.) Verdc.) production in Sub-Saharan Africa. *Agric J.* 2007;2(4):464–470.
21. Murevanhema YY, Jideani VA. Potential of Bambara groundnut (*Vigna subterranea* (L.) Verdc) milk as a probiotic beverage: A review. *Crit Rev Food Sci.* 2013;53(9):954–967. <http://dx.doi.org/10.1080/10408398.2011.574803>
22. Guillon F, Champ M. Structural and physical properties of dietary fibres, and consequences of processing on human physiology. *Food Res Int.* 2002;33:233–245. [http://dx.doi.org/10.1016/S0963-9969\(00\)00038-7](http://dx.doi.org/10.1016/S0963-9969(00)00038-7)
23. Elleuch M, Bedigian D, Roiseux O, Besbes S, Blecker C. Dietary fibre and fibre-rich by-products of food processing; Characterisation technological functionality and commercial applications: A review. *Food Chem.* 2011;124:411–421. <http://dx.doi.org/10.1016/j.foodchem.2010.06.077>
24. Huang SC, Lia TS, Cheng TC, Chan HY, Hwang SM, Hwang DF. *In vitro* interactions on glucose by different fibre materials prepared from mung bean hulls, rice bran and lemon pomace. *J Food Drug Anal.* 2009;17:307–314.
25. Rosell CM, Santos E, Collar C. Physico-chemical properties of commercial fibres from different sources: A comparative approach. *Food Res Int.* 2009;42:176–184. <http://dx.doi.org/10.1016/j.foodres.2008.10.003>
26. Tudorica CM, Kuri V, Brennan CS. Nutritional and physicochemical characteristics of dietary fiber enriched pasta. *J Agric Food Chem.* 2002;50(2):347–356. <http://dx.doi.org/10.1021/jf0106953>
27. Chau CF, Huang YL. Characterization of passion fruit seed fibres - a potential fibre source. *Food Chem.* 2004;85:189–194. <http://dx.doi.org/10.1016/j.foodchem.2003.05.009>
28. Chau CF, Huang YL, Chang FY. Effects of fibre derived from passion fruit seed on the activities of ileum mucosal enzymes and colonic bacterial enzymes in hamsters. *J Sci Food Agric.* 2005;85:2119–2124. <http://dx.doi.org/10.1002/jsfa.2230>
29. Esposito F, Arlotti G, Bonifati AM, Napolitano A, Vitale D, Fogliano V. Antioxidant activity and dietary fibre in durum wheat bran by-products. *Food Res Int.* 2005;38:1167–1173. <http://dx.doi.org/10.1016/j.foodres.2005.05.002>
30. Tosh SM, Yada S. Dietary fibres in pulse seeds and fractions: Characterisation, functional attributes and applications. *Food Res Int.* 2010;43(2):450–460. <http://dx.doi.org/10.1016/j.foodres.2009.09.005>
31. Kohajdova Z, Karovicova J, Magala M. Effect of lentil and bean flours on rheological and baking properties of wheat dough. *Chem Pap.* 2013;67(4):398–407. <http://dx.doi.org/10.2478/s11696-012-0295-3>
32. Maskus H. Pulse processing, functionality and application. Manitoba: University of Winnipeg; 2010.
33. Slavin J. Fiber and prebiotics: Mechanisms and health benefits. *Nutrients.* 2013;5(4):1417–1435. <http://dx.doi.org/10.3390/nu5041417>
34. Murevanhema YY. Evaluation of Bambara groundnut (*Vigna subterranea* (L.) Verdc) milk fermented with lactic acid bacteria as a probiotic beverage [MTech dissertation]. Cape Town: Cape Peninsula University of Technology; 2012.
35. Nti CA. Effects of Bambara groundnut (*Vigna subterranea*) variety and processing on the quality and consumer appeal for its products. *Int J Food Sci Tech.* 2009;44(11):2234–2242. <http://dx.doi.org/10.1111/j.1365-2621.2009.02064.x>
36. Sharma A. Understanding color management. USA: Thomson Delmar Learning, 2004.
37. Tomas M, Latorre G, Senti M, Marrugat J. The antioxidant function of high density lipoproteins: A new paradigm in atherosclerosis. *Rev Esp Cardiol.* 2004;57(6):557–569. [http://dx.doi.org/10.1016/s1885-5857\(06\)60630-0](http://dx.doi.org/10.1016/s1885-5857(06)60630-0)
38. Tsimikas S. Lipoproteins and oxidation. In: Bourassa MG, Tardif JC, editors. Antioxidants and cardiovascular disease. 2nd ed. Montreal: Springer; 2006. p 17. [http://dx.doi.org/10.1007/0-387-29553-4\\_2](http://dx.doi.org/10.1007/0-387-29553-4_2)
39. Betancur-Ancona D, Perza-Mercado G, Moguel-Ordóñez Y, Fuertes-Blanco S. Physicochemical characterisation of lima beans (*Phaseolus lunatus*) and jack bean (*Canavalia ensiformis*) fibrous residues. *Food Chem.* 2004;84:287–295. [http://dx.doi.org/10.1016/S0308-8146\(03\)00213-9](http://dx.doi.org/10.1016/S0308-8146(03)00213-9)
40. Saura-Calixto FD, Bravo L. Dietary fiber-associated compounds: Chemistry, analysis, and nutritional effects of polyphenols. In: Cho SS, Dreher ML, editors. Handbook of dietary fiber. New York: Marcel Dekker Inc; 2001. <http://dx.doi.org/10.1201/9780203904220.ch22>
41. Silveira JLM, Bresolin TMB. Pharmaceutical use of galactomannans. *Quim Nova.* 2011;34(2):292–299. <http://dx.doi.org/10.1590/S0100-40422011000200023>
42. Saeed F, Pasha I, Anjum FM, Sultan MT. Arabinoxylans and arabinogalactans: A comprehensive treatise. *Crit Rev Food Sci Nutr.* 2011;51(5):467–476. <http://dx.doi.org/10.1080/10408391003681418>
43. Sivam AS, Sun-Waterhouse D, Quek SY, Perera CO. Properties of bread dough with added fiber polysaccharides and phenolic antioxidants: A review. *J Food Sci.* 2010;75(8):163–174. <http://dx.doi.org/10.1111/j.1750-3841.2010.01815.x>
44. Khotimchenko Y, Khozhaenko E, Kovalev V, Khotimchenko M. Cerium binding activities of pectins isolated from the seagrasses *Zostera marina* and *Phyllospadix iwatensis*. *Mar Drugs.* 2012;10(4):834–848. <http://dx.doi.org/10.3390/md10040834>

45. Golenser J, Frankenburg S, Ehrenfreund T, Domb AJ. Efficacious treatment of experimental leishmaniasis with amphotericin B-arabinogalactan water-soluble derivatives. *Antimicrob Agents Chemother*. 1999;43(9):2209–2214.
46. Vazquez JA, Rodriguez-Amado I, Montemayor MI, Fraguas J, Gonzalez MDP, Murado MA. Chondroitin sulfate, hyaluronic acid and chitin/chitosan production using marine waste sources: Characteristics, applications and eco-friendly processes: A review. *Mar Drugs*. 2013;11(3):747–774. <http://dx.doi.org/10.3390/md11030747>
47. Dhingra D, Michael M, Rajput H, Patil RT. Dietary fibre in foods: A review. *J Food Sci Technol*. 2012;49:255–266. <http://dx.doi.org/10.1007/s13197-011-0365-5>
48. Jideani VA, Diedericks CF. Dietary fibre supplement from Bambara groundnuts (*Vigna subteranea*). South Africa – Complete patent (2014/04371); 2014 June 19. Pretoria: Companies and Intellectual Property Division; 2014



# Indicative hazard profile for strong winds in South Africa

## AUTHORS:

Andries C. Kruger<sup>1</sup>  
Dechlan L. Pillay<sup>2</sup>  
Mark van Staden<sup>2</sup>

## AFFILIATIONS:

<sup>1</sup>South African Weather Service, Pretoria, South Africa

<sup>2</sup>Early Warnings and Capability Management Systems, National Disaster Management Centre, Centurion, South Africa

## CORRESPONDENCE TO:

Andries Kruger

## EMAIL:

Andries.Kruger@weathersa.co.za

## POSTAL ADDRESS:

South African Weather Service, Private Bag X097, Pretoria 0001, South Africa

## DATES:

Received: 09 Mar. 2015

Revised: 08 July 2015

Accepted: 20 July 2015

## KEYWORDS:

wind hazard; disaster management; strong wind mechanisms; seasonal winds

## HOW TO CITE:

Kruger AC, Pillay DL, Van Staden M. Indicative hazard profile for strong winds in South Africa. *S Afr J Sci.* 2016;112(1/2), Art. #2015-0094, 11 pages. <http://dx.doi.org/10.17159/sajs.2016/20150094>

While various extreme wind studies have been undertaken for South Africa for the purpose of, amongst others, developing strong wind statistics, disaster models for the built environment and estimations of tornado risk, a general analysis of the strong wind hazard in South Africa according to the requirements of the National Disaster Management Centre is needed. The purpose of the research was to develop a national profile of the wind hazard in the country for eventual input into a national indicative risk and vulnerability profile. An analysis was undertaken with data from the South African Weather Service's long-term weather stations to quantify the wind hazard on a municipal scale, taking into account that there are more than 220 municipalities in South Africa. South Africa is influenced by various strong wind mechanisms occurring at various spatial and temporal scales. This influence is reflected in the results of the analyses which indicated that the wind hazard across South Africa is highly variable, spatially and seasonally. A general result was that the strong wind hazard is highest from the southwestern Cape towards the central and eastern parts of the Northern Cape Province, and the southeastern parts of the coast as well as the eastern interior of the Eastern Cape Province. On a seasonal basis, the southern parts of the country showed similar magnitudes of relative wind hazard throughout the year. However, further north, a strong seasonal component was evident, with lowest risk of strong winds during autumn and winter, and highest risk in spring and summer when convective activity is strongest.

## Introduction

While various extreme wind studies have been undertaken for South Africa for the purpose of, amongst others, developing statistics and disaster models for the built environment and estimation of tornado risk,<sup>1-10</sup> a general analysis of the strong wind hazard in South Africa according to the requirements of the National Disaster Management Centre (NDMC) is needed.

In South Africa, the NDMC is mandated to oversee the implementation of the *Disaster Management Act, 2002*<sup>11</sup> and the *National Disaster Management Framework 2005*<sup>12</sup> with the objective of inter alia preventing or reducing the risk of disasters, emergency preparedness, rapid and effective response to disasters and post disaster recovery.

The identification of risk related to various hazards provides the foundation for Disaster Risk Reduction activities aimed at both the prevention and mitigation of disasters and the loss of human lives. The indicative risk profile platform of the NDMC aims to provide inputs into the establishment of a national standard and subsequent guidelines in disaster risk assessments.

The NDMC approached the South African Weather Service (SAWS) to quantify a relative windstorm hazard component to form one of three components that would comprise the national indicative risk and vulnerability profile for wind storms in South Africa, in compliance with legislative requirements relating to generating indicative risk and vulnerability profiles (the other components were vulnerability and capacity). The appropriate quantification of a properly defined hazard is determined by the fact that this quantification will eventually be used in the relative assessment of risk, quantitatively defined as:

$$\text{risk} = \frac{(\text{hazard} \times \text{vulnerability})}{\text{capacity}} \quad \text{Equation 1}$$

The above equation provides a tool or measurement of relative risk, which is defined by the Disaster Risk Institute as the combination of the probability of an event and its negative consequences.<sup>13</sup>

The requirements of the project were firstly to define a windstorm hazard and retrospectively assess the conditions that meet these requirements from a historical perspective. Based on these findings, the outputs would encompass a spatial output of windstorm risk in South Africa. In addition, the seasonal quantification of the wind hazard should include the risks associated with likelihood, frequency, magnitude and predictability, as defined by the NDMC. Following the analyses of the four factors, an overall relative risk should be determined, which incorporates or considers the mentioned factors. According to the definitions of the factors, all factors can be estimated with the application of an appropriate extreme value distribution, and are therefore interlinked.

A further requirement was that the wind hazard be quantified on at least a resolution at local municipal scale. There are in excess of 200 local municipalities, and the project should allocate a relative wind hazard quantity to each of these geographical areas. While the strong wind climate would exhibit significant variability at denser spatial scales than some local municipal areas, in most cases wind measurements, on which the analyses would be based, were close to the most populated areas in the municipalities. Therefore, it was considered that the analyses of the measured data would in most cases reflect the risk to the largest part of the population (i.e. in the more urbanised areas) residing in a particular local municipality.

From the outset of the project it was agreed that the quantification of the wind hazard would be based on the statistical analyses of measured data, of which SAWS is the official custodian in South Africa. Such an approach

was also taken in the update of the extreme wind statistics of the South Africa Loading Code.<sup>8,9</sup>

## Definitions of wind hazard factors considered

The NDMC requires that 4 factors be considered in the quantification of wind hazard in South Africa. In addition, these factors, and the overall risk, should be estimated on a seasonal basis. Below are the quantifications of the factors, discussed in terms of their definitions, as agreed between SAWS and the NDMC. It should be noted that not all the definitions strictly conform to those generally found in the literature.

The *likelihood* of a wind hazard occurring can be determined by calculating the number of independent occurrences, divided by the period over which measurements were made, in years:

$$P = \frac{n}{N} \quad (n < N) \quad \text{Equation 2a}$$

$$P = 1 \quad (n \geq N) \quad \text{Equation 2b}$$

where  $P$  is the likelihood,  $n$  is the number of occurrences and  $N$  is the number of periods, e.g. years or seasons.  $P$  is limited to 1 or 100% if there is on average one wind hazard per year or season. However, it should be noted that because of the short period of data for a large fraction of utilised weather stations, as well as the high wind speeds for which the likelihood should be determined, it was advisable that the likelihoods be estimated from the application of appropriate theoretical extreme value distributions of the data.

The *frequency* of occurrence is defined as the estimated average number of days per time interval (e.g. a season or year) when the daily maximum wind gust is above the threshold which makes the occurrence of a wind hazard likely:

$$F = \frac{n}{N} \quad \text{Equation 3}$$

which is the same equation as for  $P$ , but not limited to 1 or 100% if  $n > N$ .

*Magnitude* is defined as the highest wind gust expected within a given period, e.g. a decade:

$$M - V_{\max}(N = D) \quad \text{Equation 4}$$

where  $M$  is the magnitude,  $V_{\max}$  the maximum wind speed, and  $D$  is a predetermined period, e.g. 10 years or seasons.

The NDMC links *predictability* to return periods. Predictability is measured on the number of occurrences over a given timeframe of analysis for each unit of analysis, from which the relative risk can be deduced. It runs very close to frequency; however, frequency is more focused on the number of seasonal occurrences in totality. With predictability, we would like to know that, for example, in 10 autumn seasons it was observed that in 30% of the seasons, strong wind events occurred. This is the same as the estimation of likelihood, where a percentage of the likelihood of occurrence is given:

$$T = \frac{1}{F} \quad \text{Equation 5}$$

where  $T$  is the return period, the reciprocal of the frequency.

## Determination of thresholds of hazardous wind speeds

### Definition of a wind hazard

Hazard can be defined as 'the potential occurrence of a natural or human-induced physical event that may cause loss of life, injury, or other health impacts, as well as damage and loss to property, infrastructure, livelihoods, service provision, and environmental resources'.<sup>14</sup> Wind hazards can be defined in terms of the wind speed or the origin of the

strong wind, e.g. tornadoes, tropical cyclones and thunderstorms. In terms of wind speed, the definition or threshold of a wind hazard is dependent on the socio-economic sector it relates to.

In addition to the above, the Disaster Risk Institute defines a hazard as 'a dangerous phenomenon, substance, human activity, or condition that may cause loss of life, injury, or other health impacts, property damage, loss of livelihoods and services, social and economic disruption, or environmental damage'.<sup>13</sup>

In the literature, wind hazards are most often described in terms of their origin or cause, or the danger that wind stronger than a specific threshold value might pose to a specific sector. For example, in forestry, a wind hazard is usually defined in terms of the wind speed capable of windthrow, which refers to trees uprooted or broken by wind. For the built environment, a wind hazard can be described as a wind phenomenon that exhibits wind speeds greater than the design wind speed values, or wind speeds which are in general capable of inflicting damage to structures. For the health sector, the adverse effects of strong winds on the human body are important. This has usually got to do with wind in conjunction with other factors, e.g. the potential of dust storms or the combination of strong winds and low humidity (e.g. bergwinds). For aviation, the strengths of gusts of convective origin (e.g. downburst or microbursts) are the most important. It is therefore clear that it is difficult, if not impossible, to describe or define a wind hazard holistically, as the adverse effects of strong winds are often exacerbated by the presence of other factors such as dust, low humidity and in strong or severe thunderstorms, heavy precipitation.

To complicate matters further, a wind hazard can be defined in terms of an average wind speed threshold (usually over 10 min), or gust value, which is usually assumed to be the wind speed over a period of 2–3 s, or the 'instantaneous' wind. This sometimes led to the erroneous interpretation or application of wind hazard guidelines or scales, for example in the application of the Beaufort wind scale, where wind categories are defined in terms of the average wind speed. However, examples exist where the application of the guidelines was in terms of gusts or averaged times much longer than the prescribed 10 min, e.g. 24 h.

To define a wind hazard in terms of thresholds, two options arguably exist: (1) defining a range of categories according to the most likely adverse effects it might have on a range of sectors, or (2) defining one threshold at which the wind speed will presumably have adverse effects on most of the identified sectors.

### National wind hazard warnings

As a basis, the definitions of wind hazard warnings issued by various meteorological services were reviewed. Some of the most prominent are the National Weather Service of the United States, Environment Canada, MeteoAlarm (European warning system) and the Bureau of Meteorology of Australia. The purpose of most of the alerts is to inform the public of weather conditions which might damage infrastructure and endanger lives.

For the National Weather Service of the USA, alerts of general strong wind hazards are classified into groups of two numbers on the Beaufort Scale. Warnings are given in relation to maritime conditions, land and tropical cyclones, and watches or alerts are given for specific strong wind related meteorological phenomena. These are tornadoes, severe thunderstorms and particularly dangerous thunderstorms. Significant weather advisories are issued where thunderstorms can be expected with gusts in the range of 34–50 knots (17.5–25.7 m/s). Other alerts are given where high wind speeds in conjunction with snow (blizzards), low temperatures (wind chill) and dust storms can occur.

Alerts from the Meteorological Service of Canada, under the auspices of Environment Canada, are similar to that of the USA. Watches or warnings of severe thunderstorms are given when gusts are expected to be stronger than 50 knots (25.7 m/s), as well as separate alerts for tornadoes, tropical cyclones and hurricanes. A general wind warning is issued for winds stronger than about 30 knots (15.5 m/s), gusting



at 50 knots (25.7 m/s). In addition, marine warnings are issued with different thresholds.

A study was conducted for most of the European territory to present threshold values for the warning of strong wind gusts.<sup>15</sup> This was based on the determination of return period values using validated wind gust data. Return periods indicate how extreme an event is compared to the local climate conditions and are used as a measure of the possible danger of the event and its impact on society. The choice of the new warning thresholds was based on the median of the return values for three different return periods: 1-year return periods for yellow warnings, 2-year periods for orange and five-year periods for red warnings. Modified thresholds for Norway, Ireland, the Netherlands, Germany and Spain were developed, where the need to distinguish between coastal, inland and mountainous regions was demonstrated.

In Australia, strong wind warnings are given in general, or specifically for the coastal areas, but also as part of warnings for severe thunderstorms. Threshold values for wind gusts are 90 km/h (25 m/s), and 63 km/h or 34 knots (17.5 m/s) for an average wind speed.

The South African Weather Service employs the following thresholds for its alert system: A mean wind speed of 34 knots (17.5 m/s), defined as 'gale force' according to the Beaufort scale, and a possible thunderstorm gust of 50 knots (25.7 m/s). In addition, warnings are also given for probable gusts stronger than 44 knots (22.6 m/s) along the coast. The values for mean wind speed and thunderstorm gusts correspond to the values that other meteorological services, e.g. USA National Weather Service, use for their warnings.

#### *Likely impacts of wind speeds at which warnings are issued*

The wind speed values for which official warnings are issued should be further investigated regarding their likely impact, particularly on infrastructure. However, as stated by Molarius et al.<sup>16</sup>, 'attempts to link an extreme weather event to physical consequences in a built environment using a concept of probability or likelihood are few'. Another avenue, apart from research that focuses on the likely damage to the built environment, is to use established wind scales. The most widely used is the Beaufort Wind Scale, where a mean wind speed value of 34 knots (17.5 m/s) is the lower bound of the 'gale force wind' category. This scale is used to estimate mean wind speeds according to conditions observed over land and sea. A gale force wind is described as a wind capable of breaking twigs from trees and causing cars to veer on the road. The category below gale force wind, 'near gale', is only described as being capable of inducing motion in whole trees, with effort needed to walk against the wind. From these descriptions, the gale force wind speed of 34 knots (17.5 m/s) seems to be the most appropriate lower bound for a dangerous wind warning, which is largely reflected in the warnings issued by many national meteorological services.

Regarding the warning of the likelihood of a gust speed of 50 knots (25.7 m/s), this wind speed was first used as a lower threshold in the USA for severe thunderstorm warnings. In the 1950s and 1960s, there were three types of convective watches that could be issued: tornado watches, public severe thunderstorm watches, and aviation severe thunderstorm watches. At first, for the public, the severe thunderstorm watch wind criterion was 75 mph (33.5 m/s), while the limit for aviation watches was 50 mph (22.4 m/s). Negotiations with the Air Force raised the minimum speed required for an aviation watch to 58 mph (25.9 m/s) in 1962. In 1970, the aviation severe thunderstorm and public severe thunderstorm watches were combined into a single severe thunderstorm watch, with a minimum wind gust criterion of 50 knots (25.7 m/s), to reduce confusion.<sup>17</sup> According to the risk management group, Karen Clark and Co<sup>18</sup>, wind gusts of 58 mph (25.9 m/s) can induce localised damage, as shown in Table 1. This is described as

patches of shingles missing, minor roof covering damaged, window and cladding damage to some wood frames, and some damage to unreinforced masonry and light metal structures. Significant signage damage is possible.

It is clear that the lower thresholds of mean wind speeds and gusts applied by most weather services for strong wind and severe thunderstorm alerts are appropriate in the sense that the wind speeds can be considered to be relevant to the possibility of some damage to formal infrastructure.

#### *Relevant wind speed in South Africa*

In summary, it is possible to objectively define a wind hazard as a mean wind with a strength of 34 knots (17.5 m/s) or gust of 50 knots (25.7 m/s), which with current knowledge, can be assumed to be the lower bounds of wind speeds capable of inflicting damage to infrastructure. However, in many countries, particular in developing countries like South Africa, informal settlements are still significantly contributing to the provision of shelter. These structures are of varying integrity, with the effect that it will be difficult, if not impossible, to assign a threshold wind speed at which damage is likely. A scientific investigation into the likely damage of winds of various strengths to informal settlements is needed (Goliger AM 2014, oral communication). Therefore, from the available wind damage scales, it is provisionally deduced that the lower wind speed threshold where damage to informal structures is likely, is a gust speed of about 20 m/s.

#### *Overall wind hazard*

The potential damage that wind can cause is non-linear, with the likely damage rising exponentially with wind speed, relative to the wind force, which is defined as the square of the wind speed. Therefore, it can be argued that for the overall wind hazard, the likelihood of higher wind speeds should carry progressively more weight than the lower wind speeds at which some damage to infrastructure is likely.

Table 2 summarises the Karen Clark and Co. and Beaufort wind scales, as well as meteorological warnings and the 1:50 year wind gust speed which should in most cases be used in the formal design and planning of infrastructure, according to the present South African wind loading code. It was possible to conveniently categorise wind gust speeds into five equally spaced intervals, from the lowest at 20 m/s where slight damage is likely, to 40 m/s, which is the wind speed at which extensive damage to formal infrastructure is possible. Subsequently, the overall wind hazard is defined as the sum of the likelihoods of wind speeds occurring for each of the five categories. As the likelihood is given in terms of a wind speed for a specific category or higher occurring, it follows that the likelihood of wind speeds at the higher categories carries much more weight. For example, the probability of a wind speed of 40 m/s or higher will have five times the weight of the likelihood of a wind speed of 20 m/s or higher in the overall measure for wind hazard.

Formally, we define the wind hazard as follows:

$$\text{Overall Wind Hazard} = P(V \geq 20) + P(V \geq 25) + P(V \geq 30) + P(V \geq 35) + P(V \geq 40)$$

Equation 6

where V is the wind gust value in m/s.

#### **Data**

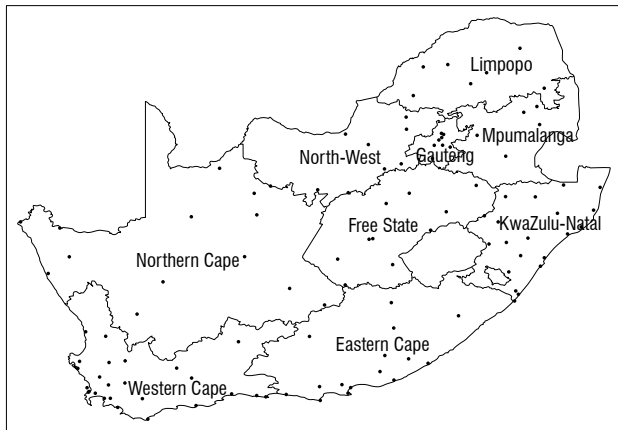
The quantification of wind hazards is based on the analysis of observed wind measurements from the SAWS network of automatic weather stations. Ten years is regarded as the minimum period from which meaningful statistical assumptions regarding the wind climate of a location can be made. The last comprehensive analysis of wind data,

**Table 1:** The Karen Clark and Co. Wind Damage Scale

None	Slight	Spotty	Localised	Structural	Failure	Extensive	Complete
0–40 mph	40–60 mph	60–80 mph	80–100 mph	100–120 mph	120–140 mph	140–160 mph	160+ mph

in 2008, utilised the data of 75 weather stations.<sup>8</sup> While a potential 94 stations were available at the time of analysis, a significant fraction of the stations had to be discarded because of inadequate exposure of the anemometers and other problems.

Currently SAWS operates 130 automatic weather stations which have been open since at least 2004, which means at least 10 years of data at the time of analysis. However, problems with data and the exposure of some stations led to the final analyses being based on the 124 weather stations presented in Figure 1.



**Figure 1:** Positions of South African Weather Service weather stations used in the wind hazard analyses.

All the stations were quality controlled according to the following procedure: All days with wind gusts stronger than 20 m/s were identified. For these days, the 5-min high resolution data of all measured weather parameters were plotted and visually inspected. This assisted in the decision to consider a wind gust as correct, or what is called a ‘spike’ in the data. Attention should inter alia be given to the difference between the mean wind speed and gust, where the ratios are generally smaller when the origin of the wind gust is at the synoptic scale compared to the mesoscale, i.e. thunderstorms. Subsequent identified erroneous wind gust values were discarded from the final data sets used for the analyses.

In the assessment by Kruger<sup>8,9</sup> it was found that the wind data from a significant fraction of SAWS weather stations were compromised by inadequate exposure of anemometers. If there is, for example, significant obstacles or nearby complex topography, it will have an influence on the representivity of the wind measurements. This is especially the case for the mean wind speed, but to a lesser degree wind gusts produced by some mesoscale weather systems, e.g. thunderstorms, where the origin of the gust is at a relatively short distance from the place

of measurement. Therefore, the representivity of the measurements had to be taken into account in the mapping of the results, especially for the regions and times of the year where wind gusts were mostly forthcoming from synoptic-scale systems. In places where the estimated wind statistics were markedly lower than for nearby stations, the exposures were investigated, e.g. by consulting photographic evidence of the surrounds and Google Earth. Where the exposures of the relevant stations were deemed to be non-standard and therefore not representative of the region, these were taken into consideration in the allocation of categories in the final mapping procedure. Eventually the aim was to produce statistics that were more or less considered to be applicable to standard conditions, which can be defined as level terrain with a roughness of approximately 0.03 m.

## Methodology

### Application of extreme value distribution

In the case of short time series, the estimation of the likelihood of strong winds should preferably be done with statistical methods particularly developed for this situation. It has been shown that the peak-over-threshold (POT) method with the fit of the exponential distribution to values above a specific threshold, produce acceptable results.<sup>13</sup>

With POT methods, all values exceeding a specific threshold are used for analysis. Usually, the generalised Pareto distribution is then fitted to the selected values and is given as:

$$F(x) = 1 - [1 - (\kappa/\alpha)(x - \xi)]^{(1/\kappa)} \quad \text{Equation 7}$$

where  $\xi$  is the selected threshold and  $\kappa$  is the shape parameter. It was shown that with data for limited periods, the safest estimation for the value of  $\kappa$  was 0.<sup>8</sup> This is consistent with Brabson and Palutikof<sup>19</sup> who, after analysing shorter and longer periods of data for the same location, came to the conclusion that the  $\kappa=0$  versions of the models make more accurate predictions of extreme wind speeds, even when a shorter period of data is utilised (in their case 13 years). Abild<sup>20</sup> came to a similar conclusion, namely that, while the generalised Pareto distribution and generalised extreme value distributions are powerful in detecting outliers, and a possible two-component population in exponential data, the tail behaviour is strongly influenced by the estimation of  $\kappa$ , and will therefore not provide reliable estimates of upper quantiles when fitted to a short record. In other words, the poor behaviour of  $\kappa$  is indicative of the data insufficiency of the short time series.

For  $\kappa = 0$ , the generalised Pareto distribution simplifies to the exponential distribution:

$$F(x) = 1 - e^{-[(x - \xi)/\alpha]} \quad \text{Equation 8}$$

**Table 2:** Wind thresholds, with indication of likely effects and damage to infrastructure

	Thresholds and likely effects				
	20 (14)	25 (18)	30 (21)	35 (25)	40 (29)
Gust (average) (m/s)	20 (14)	25 (18)	30 (21)	35 (25)	40 (29)
km/h	72	89	108	126	144
mph	45	55	67	78	89
Beaufort wind scale	7 (near gale) Effort to walk	8 (gale) Cars veer, twigs broken	9 (severe gale) Light structural damage	10 (storm) Structural damage	11 (violent storm) Widespread structural damage
Karen Clark and Co. scale	Slight damage	Slight damage	Spotty damage	Spotty damage	Localised damage
South African loading code					General prescribed 1:50 year value for design purposes
Meteorological warnings		Warning	Warning	Warning	Warning

The crossing rate of the threshold is defined as:

$$\lambda = n/M \quad \text{Equation 9}$$

where  $n$  is the total number of exceedances, and  $M$  is the total number of years of the time series.

Quantiles for specific return periods (in years) can be calculated from Abild<sup>20</sup>:

$$X_T = \xi + (\alpha/\kappa) [1 - (\lambda T)^{-\kappa}] \quad \kappa \neq 0 \quad \text{Equation 10}$$

$$X_T = \xi + \alpha \ln(\lambda T) \quad \kappa = 0 \quad \text{Equation 11}$$

The distribution parameters  $\alpha$  and  $k$  can be estimated with:

$$\hat{\kappa} = [b_0 / (2b_1 - b_0)] - 2 \quad \text{Equation 12}$$

$$\hat{\alpha} = (1 + \hat{\kappa}) b_0 \quad \text{Equation 13}$$

which are valid within the range  $-0.5 < \kappa < 0.5$ .

As with the other methods, selected extreme values should be independent. To ensure independence, choices have to be made between the threshold value and the separation time between the selected values. The choices of the separation time vary between authors, with e.g. Cook<sup>21</sup> and Gusella<sup>22</sup> using 48 h for European wind climates and Walshaw<sup>23</sup> using 60 h for Sheffield data. The same applies for the selection of the threshold, where Cook<sup>21</sup> decided on a threshold of 15 m/s for hourly mean wind speeds and 30 m/s for wind gusts, while Brabson and Palutikof<sup>24</sup> found that a threshold of 32–36 m/s is most suitable for Shetland in the north of the United Kingdom. It therefore becomes clear that the optimum choices of threshold and separation time depend on location and specifically the local wind climate. To aid in the threshold selection, various techniques exist, such as the conditional mean exceedance (CME) graphs, which is a plot of the mean excess over threshold as a function of the threshold value.<sup>24,25</sup> Brabson and Palutikof<sup>19</sup> defined an independent event index:

$$\varepsilon = (\text{independent events}) / (\text{total events}) \quad \text{Equation 14}$$

and calculated the value of  $\varepsilon$  for various combinations of threshold and 'dead time', i.e. the minimum period between selected values. An independent event can be defined as an event where the cause is different from other events before and after the particular event. The ideal is for  $\varepsilon$  to be as close as possible to unity, but a value of  $\varepsilon=0.8$  is sufficient to obtain accurate quantile estimates with the POT method.

#### Estimation of four factors using the peak-over-threshold method

All four factors discussed can be estimated with the application of the POT method. From Equations 4, 5 and 6:

$$T = \frac{1}{\lambda} * e^{\left(\frac{x-\xi}{\alpha}\right)} \quad \text{Equation 15}$$

where  $T$  is the return period.

The *predictability* is simply the application of Equation 9 with  $x=20$  m/s.

The *frequency*, or estimated number of days per season with gusts higher than 20, is given by  $1/T$ , if  $x=20$  m/s.

For *probability*, if the days per season are equal to or greater than 1, then the probability is 100%, else the probability is given as the total number of days in a season multiplied by  $1/T$ .

The *magnitude* is calculated directly from Equation 7, with  $T=10$  years.

#### Relative categorisation of overall wind hazard

The wind hazard across South Africa should be relatively quantified on a scale with five categories. To achieve this, the probabilities of occurrence of wind gusts in the five categories presented in Table 2 were utilised on a linearly weighted scale. This was done by estimating, per season and per station, the probabilities of wind gusts stronger than the lower margins and calculating the sum of the five probabilities. The range of results was then divided into five equal categories, assuming that the sum of probabilities increased linearly. Figure 2 presents the sums of probabilities in increasing order, as well as the least-squares fitting of a straight line on the curve. One can see that the assumption of a linear increase in value vs rank can be accepted, with a statistically significant correlation factor of 0.96. It is only at the higher ranks that there is a very marked deviation from the straight line.

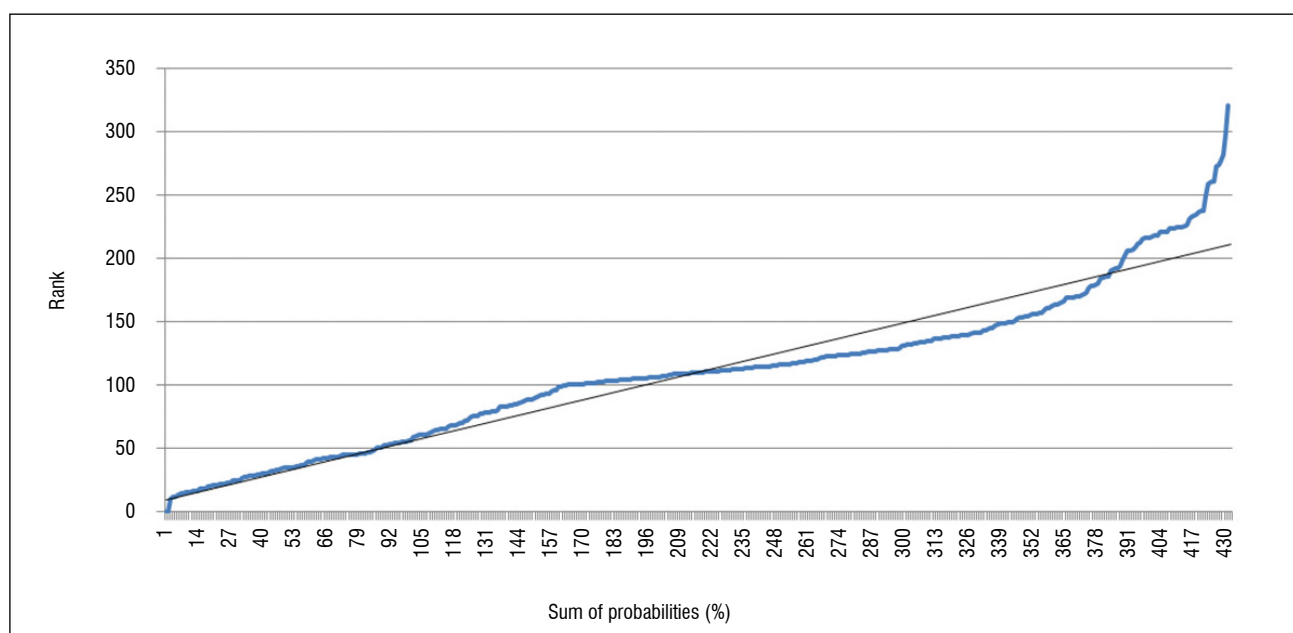


Figure 2: Sums of probabilities in increasing order vs rank for all stations and seasons combined. The straight line represents the least-squares linear fit between the rank and the sum of probabilities.

## Results

The NDMC required results on a seasonal basis for the four factors as well as the relative overall wind hazard. In addition, an annual map of the relative wind hazard was required for eventual application to the estimation of the risk of a wind hazard.

### Likelihood

Figure 3 presents the likelihood of a wind hazard, defined by the occurrence of a gust equal or higher than 20 m/s per season. In general, most of South Africa exhibits a high probability of at least one incident of a wind gust of 20 m/s or higher, except for the northern and northeastern parts during autumn and winter. This is mostly because of the diminished probability of strong thunderstorms occurring in the summer rainfall regions during those seasons. In the central and southern parts of the country, there is a likelihood for strong thunderstorms during the summer months, as well as strong winds from cold fronts during winter, when strong northwesterly winds often occur before the passage of a cold front.<sup>6</sup>

### Frequency

Figure 4 presents the frequency, defined as the estimated average number of days per season when the daily maximum wind gust is above 20 m/s. During the summer months, the highest occurrences of strong wind gusts are estimated for the southwestern Cape, the Nelson Mandela metropole and the central Karoo, at 5 days or more per season. The northern and northeastern parts also show their highest frequency in

summer because of strong thunderstorms, although only for about 1 or 2 days per season.

In autumn, the high frequencies in the southern and central regions are almost the same as in summer. However, in winter the Western Cape shows an increased frequency, resulting from the increased likelihood of strong frontal systems. This situation continues into spring, where the prevalence of strong southeasterlies increases over the southwestern parts. In the northern parts, the frequencies of strong wind gusts return to the summer situation.

### Magnitude

Figure 5 presents the magnitude per season, which is defined as the highest wind gust expected in 10 years. In summer, the highest wind gusts probable over a decadal period are in the central interior, where De Aar and Kimberley can expect wind gusts in excess of 35 m/s. A wind gust of this magnitude equates to a storm where some structural damage to the built environment is highly likely.

The regional situation in summer continued into autumn, although with lower magnitude. In winter the magnitude increased over the south and southern interior, where Worcester and Beaufort West could expect wind gust magnitudes higher than 35 m/s per decade. The spring situation was similar to autumn.

In the northern and northeastern parts of the country, the seasonal magnitudes vary between 20 and 25 m/s in spring and summer, to lower than 20 m/s in many places in the far north during autumn and winter.

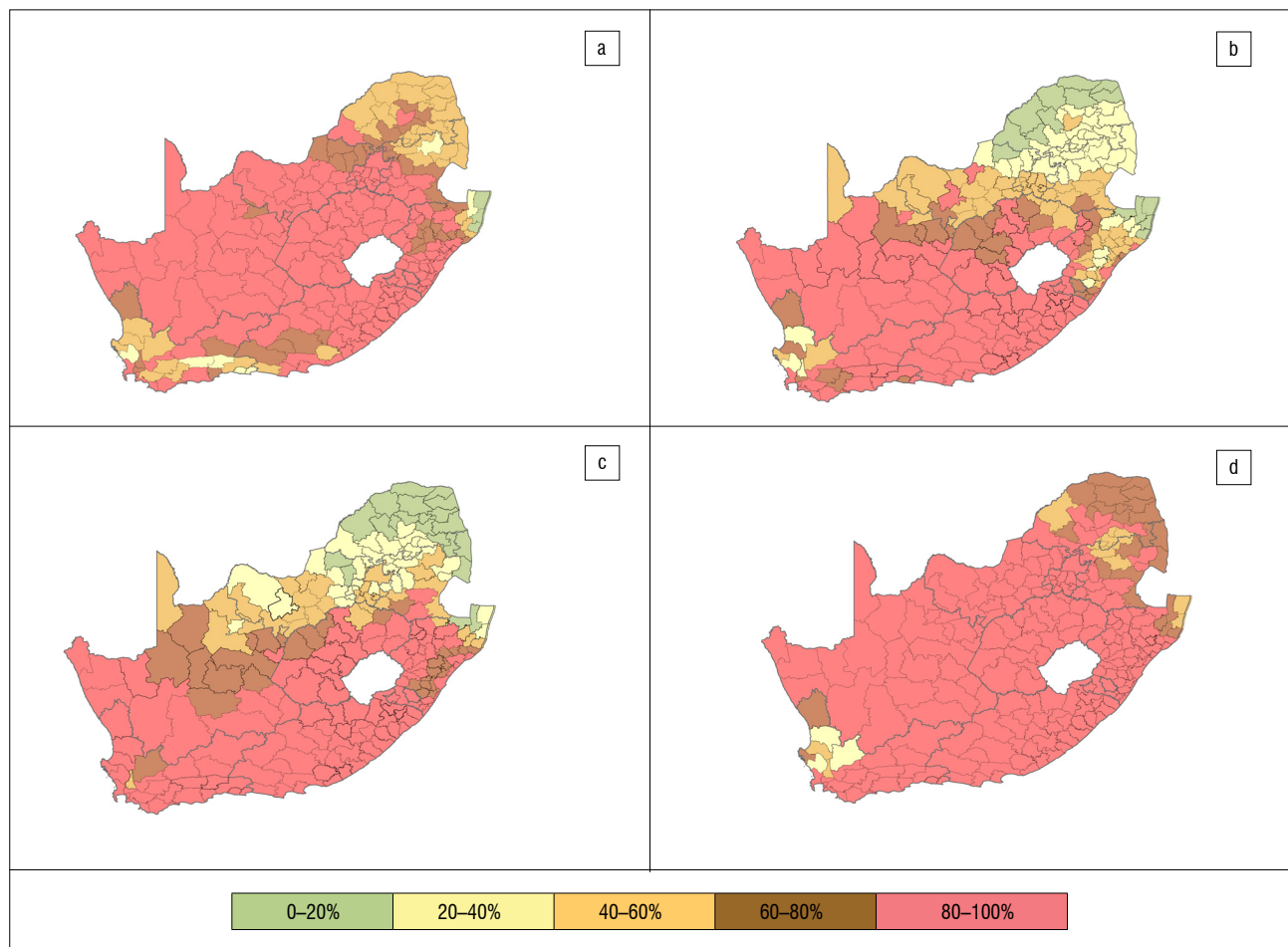
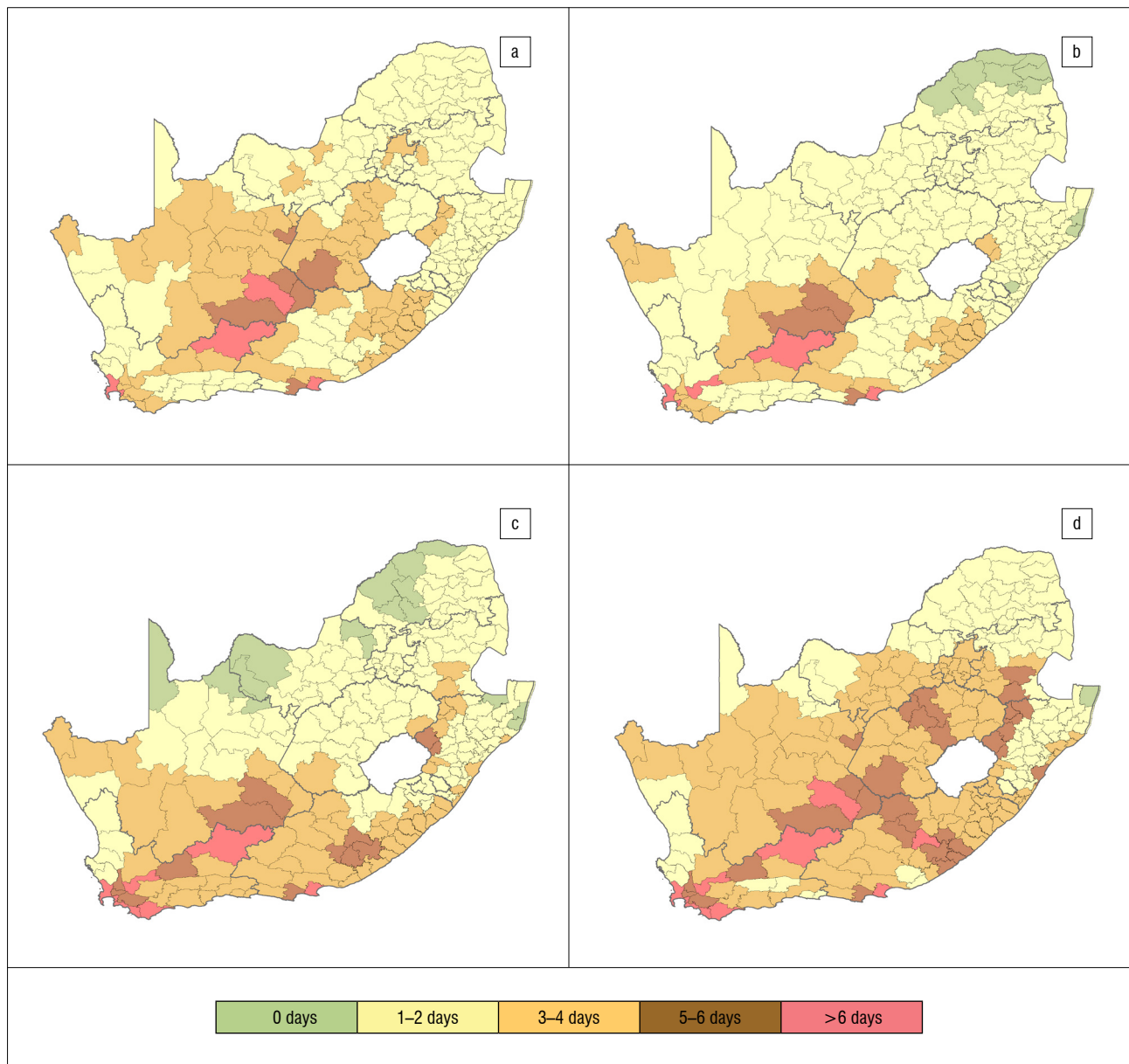


Figure 3: Likelihood of a wind hazard, defined by the probability of occurrence of a gust equal to or greater than 20 m/s, per season: (a) Summer (DJF), (b) autumn (MAM), winter (JJA) and (d) spring (SON).





**Figure 4:** Frequency of wind hazard per season, defined as the estimated average number of days per season when the daily maximum wind gust is above 20 m/s: (a) Summer (DJF), (b) autumn (MAM), winter (JJA) and (d) spring (SON).

### Predictability

Figure 6 presents the predictability per season, which is defined as the estimated return period of a wind gust of 20 m/s or higher. In summer, most of South Africa from Gauteng southwards can expect wind gusts higher than 20 m/s. This situation also applies to spring, where it can be expected that many places north of Gauteng will experience strong wind gusts at least once per season.

In the autumn and winter months, the expectations of strong wind gusts in the northeast diminish, with return periods of 2–3 years or longer expected for strong wind gusts to occur. However, in the southwestern Cape there is a marked increase in the expectation of strong wind gusts in winter, at least once per season.

### Overall wind hazard

It was shown that all four factors could be estimated with the application of an appropriate extreme value distribution. It follows that the results of the determination of the factors are therefore strongly linked

quantitatively. Similar regions and seasons should therefore indicate relatively high likelihoods and frequencies of strong wind gusts, as reflected in the maps for the four factors considered.

For the estimation of risk, a relative indicator of the wind hazard was required. The NDMC requires five relative wind hazard classes or categories, with more or less equal spatial extents over South Africa. The previous discussions of the results of the four factors provided further motivation that a measure of the overall wind hazard should be based on the probabilities of strong wind gusts.

In summary, the seasonal results of the overall relative wind hazard presented in Figure 7 show the following: A larger section of South Africa is subjected to very strong winds in summer and spring than during the other seasons of the year. In the south and southeast, the higher hazard categories are, as expected, more prevalent during winter and spring. The central parts, e.g. central and southern Free State and northeastern Karoo experience high wind hazards during summer and spring, when

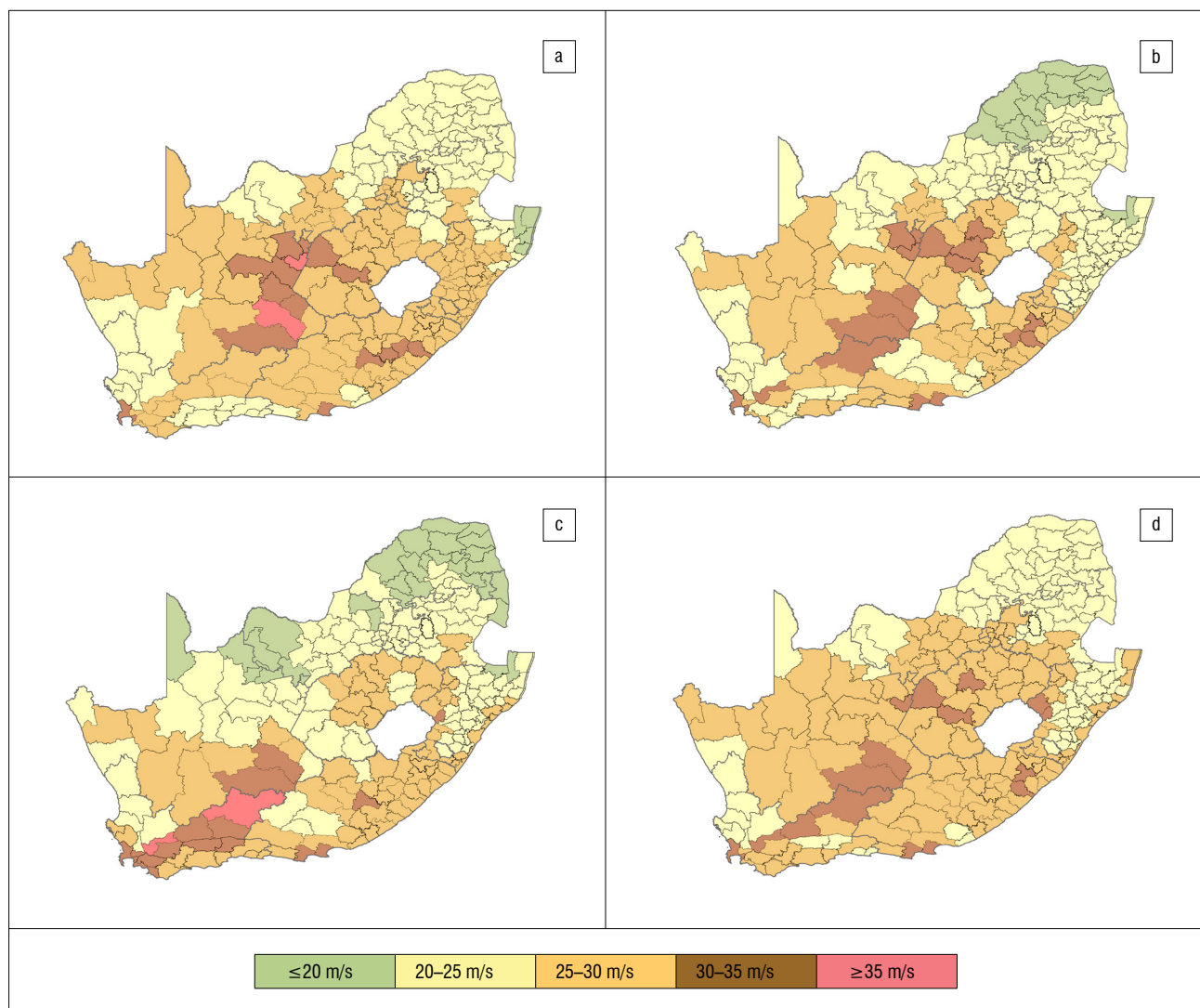
strong northeasterly winds associated with troughs over the interior are prevalent, together with the increased likelihood of strong thunderstorms.

The northern and northeastern parts of the country show the same tendency toward higher wind hazards in summer and spring as in the central parts. However, the wind hazard in this region remains relatively low, with a highest category of 3 attained in some places during spring.

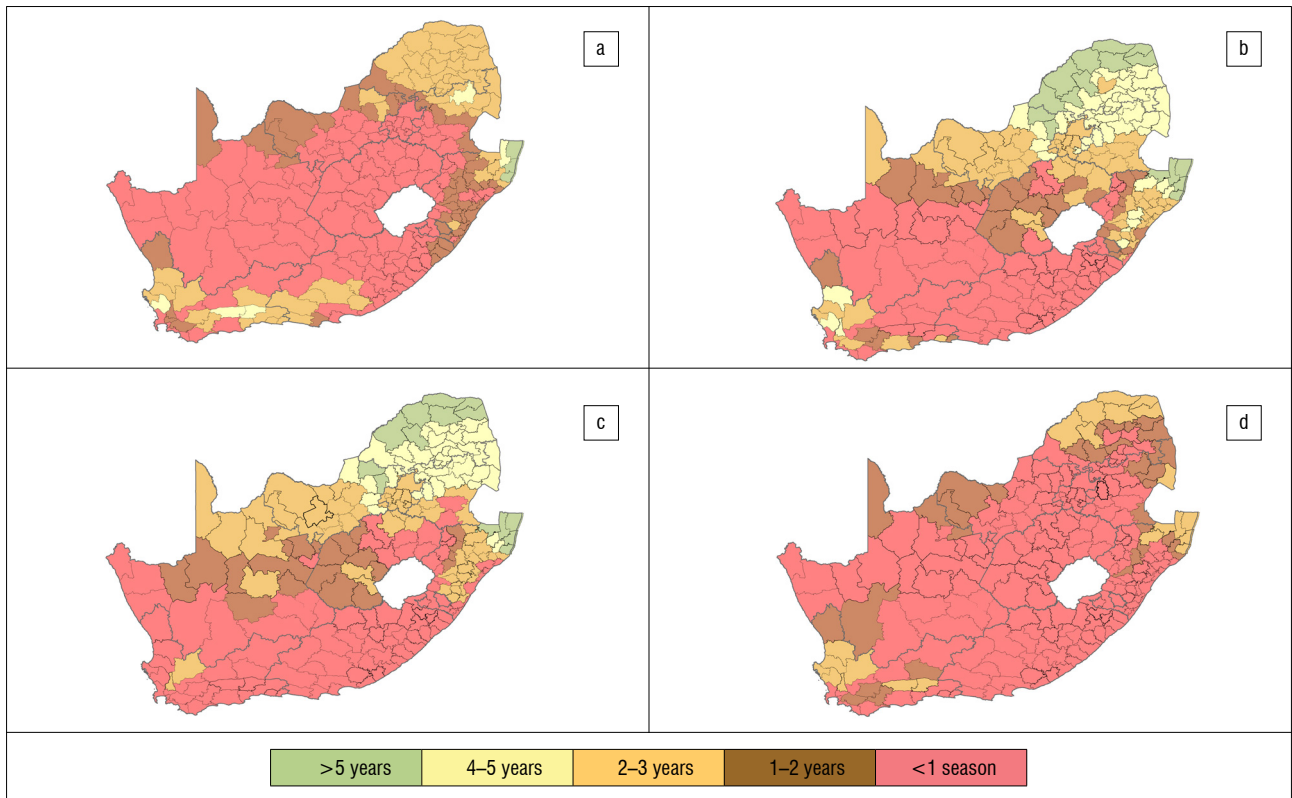
The overall all-year relative wind hazard risk map, presented in Figure 8, provides a summary of the four seasons combined. The highest relative wind hazard is in:

- Southwestern Cape, further northeastwards from Worcester, Beaufort West, De Aar, Kimberley/Bloemfontein and Welkom in the north, (mainly because of synoptic systems in the southwestern Cape, i.e. cold fronts and ridging of the Atlantic high pressure system, and central interior, i.e. deep troughs, and thunderstorms in the north).
- Nelson Mandela metropole and surrounds (because of synoptic scale systems, i.e. cold fronts)
- Parts of the escarpment in the northeast, e.g. Van Reenen to Ladysmith, Bergville and Newcastle (because of synoptic-scale systems, i.e. cold fronts).
- Eastern half of the Eastern Cape (former Transkei).

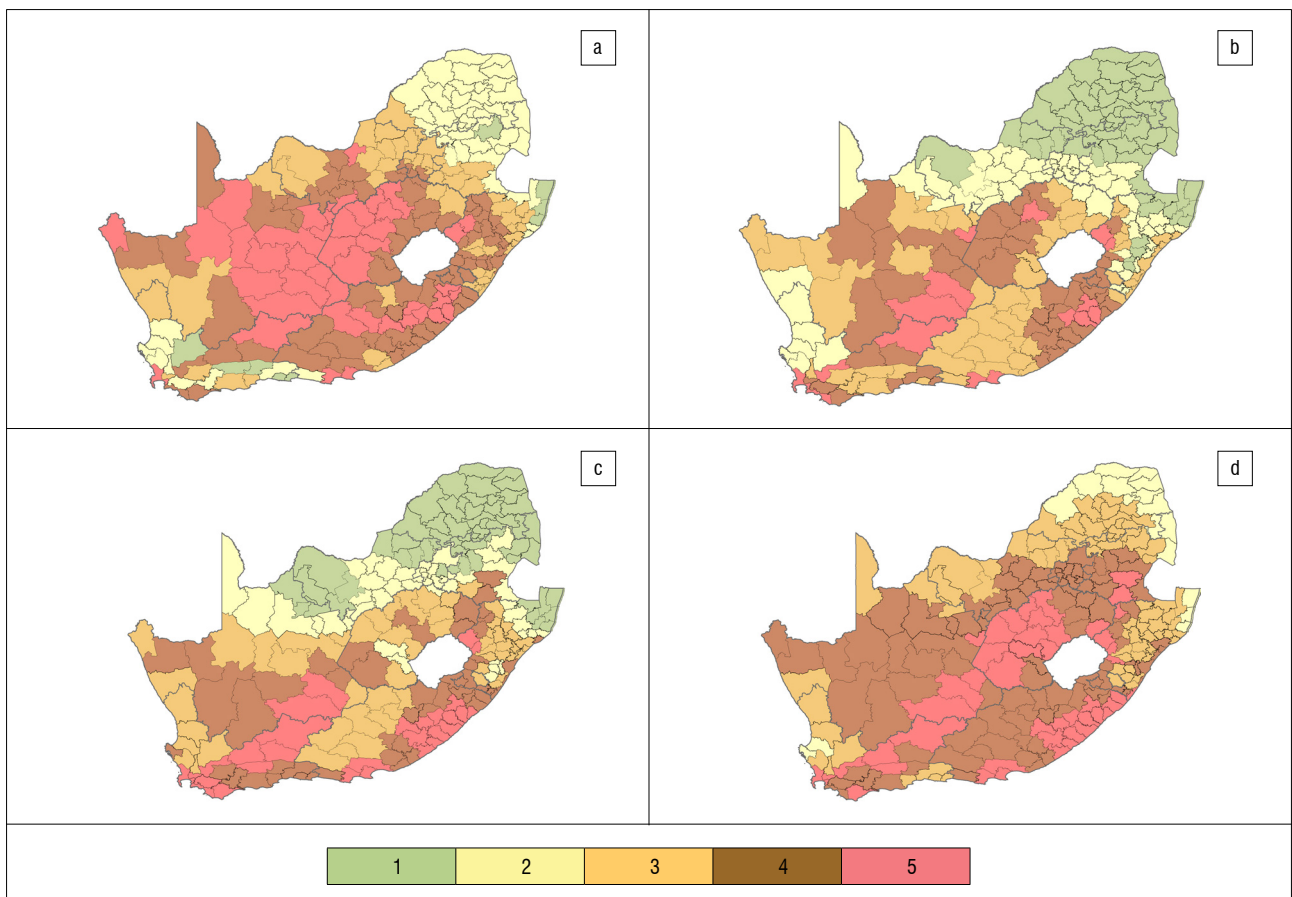
It should be noted that by far most of the disastrous implications of extreme wind events in South Africa (in terms of material and human loss) were as a consequence of tornadic events. Because of their limited spatial extent and low frequency of occurrence, these events do not get recorded by a measurement network such as that of SAWS, on which this research is based. The outputs of comprehensive research in this regard have been published by the Council of Industrial and Scientific Research (CSIR) and the then South African Weather Bureau.<sup>1</sup> It is therefore important to compare the results of the tornado study with the all-year relative wind hazard risk map presented in Figure 8. Strong tornadoes which are capable of inflicting extensive damage most frequently occur in eastern half of South Africa, particularly over the southern part of the KwaZulu-Natal province interior and Gauteng province (Figure 9). In the overall wind hazard map in Figure 8, these regions are shown to be at moderate (3/5) to below-moderate risk (2–3/5) of strong winds. While the probability of occurrence for a tornado at a single point in these regions is extremely low (about 0.0001 occurrences per year), it is still important to note, especially in the case of designing extensive infrastructure, such as power lines, that extreme weather events with a very low probability of occurrence (e.g. tornadoes and tropical cyclones) are not fully reflected in the final results of this research. Here the probability of wind speeds higher than 20 m/s is reflected in the output, which is not of a similar order of magnitude to wind speeds from tornadoes.



**Figure 5:** Magnitude, defined as the highest wind gust in m/s to expect in 10 years, per season: (a) Summer (DJF), (b) autumn (MAM), (c) winter (JJA) and (d) spring (SON).



**Figure 6:** Predictability per season, defined as the estimated return period in years of a wind gust of 20 m/s or higher: (a) Summer (DJF), (b) autumn (MAM), (c) winter (JJA) and (d) spring (SON).



**Figure 7:** Relative wind hazard per season: (a) Summer (DJF), (b) autumn (MAM), (c) winter (JJA) and (d) spring (SON).



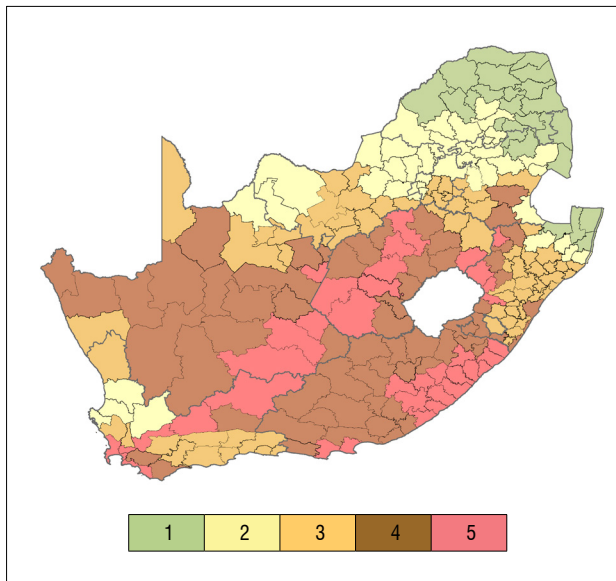


Figure 8: Relative wind hazard for all seasons combined on a scale of 1 (low wind hazard risk) to 5 (highest risk).

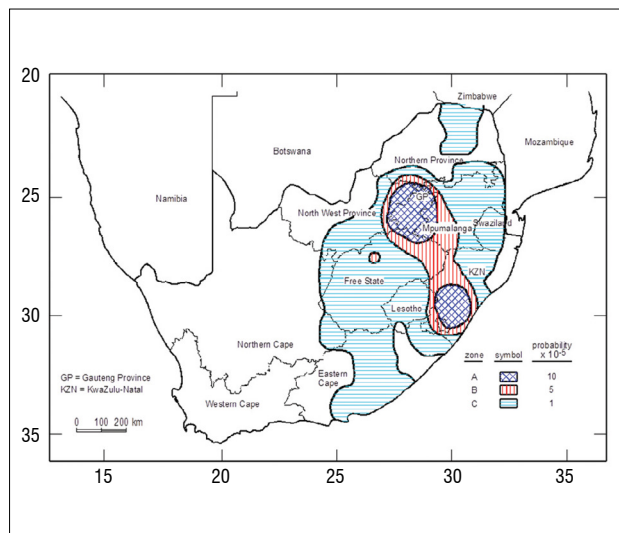


Figure 9: Mean rate of occurrence of tornadoes per annum, excluding tornadoes of intensity FO as analysed by Goliger et al.<sup>1</sup>

## Conclusions

From the literature and information of relevant built environment expertise, the definition of a windstorm hazard, in terms of magnitude, is sector dependent. Also, as is the case for the built environment, the definition of wind hazard can depend on the resilience of the socio-economic sector. Therefore, the determination of a specific lower wind speed threshold for the analysis required a measure of subjectivity regarding the approximate value at which damage to informal infrastructure is possible. More objectivity in the development of such a definition should be possible in developed countries, where the utilisation of informal structures for shelter is low. From available wind scales, the wind speed at which damage is likely was estimated at a wind gust of about 20 m/s. This threshold was therefore used in the overall definition of a wind hazard, and also in the definitions of the four parameters defined by the disaster management sector.

The seasonal quantification of the windstorm hazard in terms of the parameters likelihood, frequency, magnitude and predictability, showed clear seasonal variation in windstorm hazard. The relative hazard was for all parameters the highest in the southern parts of South Africa for the autumn and winter months, expanding northwards during spring and

summer. The same shifting patterns were found for all of the mentioned parameters, as well as in the analysis of overall relative risk.

Ultimately, it could be shown that some regions in the country are, in general, more prone to wind hazards than others at the annual temporal scale, with the regions at highest risk being in the south and southeast and along parts of the escarpment. Strong wind events with very low probability, e.g. tornadoes and tropical cyclones, were not considered. It is impossible to consider these events in the forward planning to mitigate against strong wind disasters, because of the rarity of occurrence at specific point locations. In contrast, it could be possible for countries with regions prone to the occurrence of tornadoes and tropical cyclones, e.g. the United States midwest, to take these events into account in a quantification of strong wind hazard.

The outputs of the research will serve as essential input in the determination of wind hazard risk (with the additional two inputs, vulnerability and capacity) by the NDMC.

## Acknowledgements

We acknowledge the NDMC for funding the project and two anonymous reviewers whose inputs and comments significantly improved the contents.

## Authors' contributions

A.C.K. analysed the climate data, developed and performed relevant statistical analysis, developed the end products in GIS and authored most of the submitted manuscript. D.L.P. initiated the research, contributed to the design of the parameters of risk included, and ensured that the end products of the research fulfilled the requirements from a disaster risk reduction perspective. He contributed to the manuscript by revision to ensure that the relevance of the work to the NDMC was well-communicated. M.v.S. provided technical perspective on the end product required by NDMC, provided background to the significance of the results, and played the main role in the development of the factors considered in the design of the hazard profile to fulfil the requirements for the use of the wind hazard profile in the development of the subsequent risk profile. He also contributed to the writing of the manuscript.

## References

- Goliger AM, Milford RV, Adam BF, Edwards M. *Inkanyamba: Tornadoes in South Africa*. Pretoria: CSIR and South African Weather Bureau; 1997.
- Goliger AM. *Development of a wind damage and disaster risk model for South Africa* [PhD thesis]. Stellenbosch: Stellenbosch University; 2002.
- Goliger AM, Retief JV, Niemann H-J. Wind damage/disaster model. Part 1: Generic algorithm. In: *Proceedings of the 11th International Conference on Wind Engineering (ICWE)*; 2003 June 2–5; Lubbock, TX, USA. Lubbock, TX: ICWE; 2003. p. 1023–1030.
- Goliger AM, Retief JV, Niemann H-J. Wind damage/disaster model. Part 1: Application. In: *Proceedings of the 11th International Conference on Wind Engineering (ICWE)*; 2003 June 2–5; Lubbock, TX, USA. Lubbock, TX: ICWE; 2003. p. 1031–1038.
- Goliger AM, Retief JV, Niemann H-J. A development process of a wind disaster risk model for South Africa. In: *Proceedings of the 2nd International Conference on Structural Engineering, Mechanics and Computation*; 2004 July; Cape Town, South Africa. Cape Town: University of Cape Town; 2004 p. 843–848.
- Goliger AM, Retief JV. Severe wind phenomena in southern Africa and the related damage. *J Wind Eng Ind Aerod*. 2007;95:1065–1078. <http://dx.doi.org/10.1016/j.jweia.2007.01.029>
- Goliger AM, Mahachi J. Effects of severe windstorms on housing stock in South Africa – Adequate constructions vs extreme weather events. Paper presented at: 11th International Housing and Home Warranty Conference (IHHWC08); 2008 September 6–10; Edinburgh, UK.
- Kruger AC, Retief JV, Goliger AM. Strong winds in South Africa: Part I – Application of estimation methods. *J S Afr Inst Civ Eng*. 2013;2:29–45.
- Kruger AC, Retief JV, Goliger AM. Strong winds in South Africa: Part II – Mapping of updated statistics. *J S Afr Inst Civ Eng*. 2013;2:46–58.



10. Kruger AC, Goliger AM, Retief JV, Sekele SS. Clustering of extreme winds in the mixed climate of South Africa. *Wind Struct.* 2012;15(2):87–109. <http://dx.doi.org/10.12989/was.2012.15.2.087>
11. Disaster Management Act, 2002 (Act 57 of 2002) [document on the Internet]. c2003 [cited 2014 December 11]. Available from: <http://www.cogta.gov.za/sites/cogtapub/Acts/DISASTER%20MANAGEMENT%20ACT.pdf>
12. National Disaster Management Framework. Government Notice 654 of April 2005 [document on the Internet]. c2005 [cited 2014 December 11]. Available from: <http://www.ndmc.gov.za/portals/0/WebDocuments/Guidelines/DM/downloads/SJ%208%20NDMF%202005.pdf>
13. The United Nations International strategy for disaster reduction (UNISDR). UNISDR terminology [document on the Internet]. c2009 [cited 2015 January 16]. Available from: [http://www.unisdr.org/files/7817\\_UNISDRterminologyEnglish.pdf](http://www.unisdr.org/files/7817_UNISDRterminologyEnglish.pdf)
14. Intergovernmental Panel on Climate Change. Impacts, adaptation, and vulnerability: Summary for policymakers. Geneva: IPCC; 2014. Available from: [https://ipcc-wg2.gov/AR5/images/uploads/IPCC\\_WG2AR5\\_SPM\\_Approved.pdf](https://ipcc-wg2.gov/AR5/images/uploads/IPCC_WG2AR5_SPM_Approved.pdf)
15. Stepek A, Wijnant IL, Van der Schrier G, Van den Besselaar EJM, Klein Tank AMG. Severe wind gust thresholds for Meteoalarm derived from uniform return periods in ECA&D [document on the Internet]. c2010 [cited 2015 January 17]. Available from: [http://eca.knmi.nl/documents/Natural\\_Hazards\\_versie6.pdf](http://eca.knmi.nl/documents/Natural_Hazards_versie6.pdf)
16. Molarius R, Könönen V, Leviäkangas P, Zulkarnain, Rönty J, Hietajärvi A-M, Oiva K. The extreme weather risk indicators (EWRI) for the European transport system. *Nat Hazards.* 2014;2:189–210. <http://dx.doi.org/10.1007/s11069-013-0650-x>
17. Severe Weather 101. Frequently asked questions about damaging winds [document on the Internet]. No date [cited 2015 January 18]. Available from: <https://www.nssl.noaa.gov/education/svrwx101/wind/faq/>
18. Karen Clark and Co [document on the Internet]. c2014 [cited 2015 January 17]. Available from: <http://www.karenclarkandco.com>
19. Brabson BB, Palutikof JP. Tests of the generalized Pareto distribution for predicting extreme wind speeds. *J Appl Meteorol.* 2000;39:1627–1640. [http://dx.doi.org/10.1175/1520-0450\(2000\)039<1627:TOTGPD>2.0.CO;2](http://dx.doi.org/10.1175/1520-0450(2000)039<1627:TOTGPD>2.0.CO;2)
20. Abild J. Application of the wind atlas method to extremes of wind climatology. Technical report Risoe-R-722 (EN). Roskilde, Denmark: Risø National Laboratory; 1994.
21. Cook NJ. Towards better estimation of wind speeds. *J Wind Eng Ind Aerod.* 1982;9:295–323. [http://dx.doi.org/10.1016/0167-6105\(82\)90021-6](http://dx.doi.org/10.1016/0167-6105(82)90021-6)
22. Gusella V. Estimation of extreme winds from short-term records. *J Struct Eng.* 1991;117:375–390. [http://dx.doi.org/10.1061/\(ASCE\)0733-9445\(1991\)117:2\(375\)](http://dx.doi.org/10.1061/(ASCE)0733-9445(1991)117:2(375))
23. Walshaw D. Getting the most from your extreme wind data: A step by step guide. *J Res Natl Inst Stand Technol.* 1994;99:399–411. <http://dx.doi.org/10.6028/jres.099.038>
24. Davison AC. Modelling excesses over high thresholds, with an application. In: Tiago de Oliveira J, editor. *Statistical extremes and applications*. Dordrecht: Springer; 1984. p. 461–482. [http://dx.doi.org/10.1007/978-94-017-3069-3\\_34](http://dx.doi.org/10.1007/978-94-017-3069-3_34)
25. Ledermann W, Lloyd E, Vajda S, Alexander C, editors. *Handbook of applicable mathematics volume 7: Supplement*. Chichester: Wiley-Interscience; 1990.



# Radiological and genetic analysis of a Late Iron Age mummy from the Tuli Block, Botswana

## AUTHORS:

Frank J. Rühli<sup>1</sup>

Maryna Steyn<sup>2,3</sup>

Morongwa N. Mosothwane<sup>2,4</sup>

Lena Öhrström<sup>1</sup>

Molebogeng K. Bodiba<sup>2</sup>

Abigail Bouwman<sup>1</sup>

## AFFILIATIONS:

<sup>1</sup>Institute of Evolutionary Medicine, University of Zurich, Zurich, Switzerland

<sup>2</sup>Forensic Anthropology Research Centre, Department of Anatomy, University of Pretoria, Pretoria, South Africa

<sup>3</sup>School of Anatomical Sciences, University of the Witwatersrand, Johannesburg, South Africa

<sup>4</sup>Archaeology Unit, University of Botswana, Gaborone, Botswana

## CORRESPONDENCE TO:

Maryna Steyn

## EMAIL:

maryna.steyn@wits.ac.za

## POSTAL ADDRESS:

School of Anatomical Sciences,  
7 York Road, Parktown 2193,  
South Africa

## DATES:

Received: 13 Apr. 2015

Revised: 14 July 2015

Accepted: 24 July 2015

## KEYWORDS:

aDNA; computerised tomography; mummified remains; population relationships; southern Africa

## HOW TO CITE:

Rühli FJ, Steyn M, Mosothwane MN, Öhrström L, Bodiba MK, Bouwman A. Radiological and genetic analysis of a Late Iron Age mummy from the Tuli Block, Botswana. *S Afr J Sci.* 2016;112(1/2), Art. #2015-0139, 7 pages. <http://dx.doi.org/10.17159/sajs.2016/20150139>

Mummified human remains are valuable sources of information on past populations. Here we report on the radiological and molecular findings of a partially mummified individual found in northern Botswana. This desiccated mummy from the Tuli region is the first to have been reported from this region. The remains were those of an older male adult of African origin. He was interred in a tightly flexed position and wrapped in an animal skin. Computerised tomography (CT) scanning revealed that none of the internal organs was preserved. Multiple post-mortem alterations are seen, but apart from some degenerative changes of the lower vertebral column, the axial skeleton has remained intact. The advanced osteophytosis suggests an older age than what was previously estimated. The aDNA analysis confirms Sotho–Tswana and possibly Khoesan genetic relatedness, as could be expected from individuals from that region. These results represent one of the first CT scans of a mummified individual from southern Africa, and also the first successful aDNA extraction from such remains.

## Introduction

Mummified remains are uncommon finds, especially in southern Africa. They are valuable sources of information on past populations<sup>1–5</sup>, and they also tend to evoke considerable emotion as they provide a very vivid view of life in the past<sup>6</sup>. Such remains in southern Africa are usually preserved as a result of desiccation, and they are thus found in caves, rock shelters or other areas where they are protected from water. Many other factors also contribute to mummification, such as level of humidity, temperature, and covering or clothing. For example, mummified remains have been recovered from a rock shelter in the Kouga region<sup>7</sup>, the Historic Cave in the Makapan Valley<sup>8</sup> and Eland Cave<sup>8</sup>.

Recently, Mosothwane<sup>9</sup> described the discovery of a partially mummified individual from the Tuli region of Botswana – the first such remains found in Botswana (Figure 1). The site is located in a privately owned game lodge, the Northern Tuli Game Reserve, approximately 20 km west of the Shashe-Limpopo confluence. Bone collagen from the mummy was radiocarbon dated to  $140 \pm 30$  BP or AD 1675–1735 (2-sigma calibrated) or alternatively from AD 1800 to post-1950 (Beta Analytic; 400348).



Figure 1: The intact mummy, covered with animal skin, discovered in the Tuli region of Botswana.

The remains were found in a shallow grave at the base of a cliff. The cliff provided a roof directly above the grave which then limited the amount of rain water falling directly on the grave. The cliff also provided full-day shade coverage above the grave. The body was tightly flexed at the hips and elbows and wrapped with a cow skin with the furry part against the human remains. It was then tied with a rope made of plant fibre.

The remains were nearly complete, except for parts of the right femur and some phalanges which may have been carried off by scavenging animals. Skin, tendons, hair and nails were preserved. Based on the facial hair and skeletal characteristics that were visible, the remains were assessed to be that of a male adult aged between 40 and 55 years. When the animal skin was removed as part of the detailed study described below, soft tissues of the genital region could be visualised, further confirming the sex of the individual. The initial publication by Mosothwane<sup>9</sup> only provided a basic morphological description of the remains, with no detailed or specialised analyses.

Specialised medical analysis on mummified remains including diagnostic imaging technologies such as X-ray or computerised tomography (CT)<sup>10</sup>, isotopic analyses<sup>11,12</sup>, DNA extraction<sup>13-15</sup> and histological analyses<sup>16</sup> have become the standard for the assessment of mummies across the world and to a limited extent also in southern Africa<sup>6,7</sup>. Such studies allow us to answer more questions on the origins, diseases, and reasons for preservation of the remains.

The aim of this paper is to report on the outcomes of specialised analyses of the Tuli mummy, including CT scanning and aDNA analysis.

## Materials and methods

The Tuli mummy is currently housed at the Botswana National Museum in Gaborone, Botswana. The mummified remains were radiographically and macroscopically analysed. In addition, various samples were extracted. These samples included those for aDNA extraction from soft tissue and teeth and radiocarbon dating from bone (collagen). aDNA comparative samples were obtained from the GenBank database. Overall the macroscopic sample preservation was extremely good as indicated by the results of the radiographic and osteometric analysis. Likewise, bone collagen preservation was good and yielded reliable radiocarbon dates.

The remains of the Tuli individual were examined using a clinical CT scanner (Toshiba) at the Bokamoso Private Hospital in Gaborone in August 2011. Imaging parameters were as follows: 512×512 matrix; 5 mm slice thickness, 80 mA tube current, 120 kV tube voltage and 1.087×1.087 pixel spacing. In total, three series were obtained: axial (194 images), coronal (38 images) and sagittal (30 images). The data sets were processed with OsiriX-64 bit (version 5.8.5) software, including multiplanar reconstructions and three-dimensional volume rendering.

After the remains were scanned, the cow hide was carefully removed and several samples were collected at the archaeology laboratory of the University of Botswana. This sampling was done under controlled circumstances, with investigators wearing masks and gloves so as to avoid contaminating the remains. However, it should be taken into account that the remains were not excavated under such controlled conditions, but it can be assumed that the cow hide would have to some extent protected the remains from contamination. These samples were exported and DNA extraction was undertaken at dedicated aDNA laboratories in Zurich (Switzerland) at the Institute for Evolutionary Medicine. Other samples were sent to Beta Analytic Inc. (Miami, FL, USA) for radiocarbon dating.

Both bone and soft tissue were used to extract DNA and this extraction was performed under strict, sterile conditions. A modified phenol-chloroform method<sup>17,18</sup> was used, in which elution of the DNA was done by using a column tube with a chaotropic agent. DNA concentration was measured using a digital spectrophotometer (Qubit). This method was followed by standard polymerase chain reaction (PCR) amplification of the *SRY* gene and real-time PCR for the *AMELX* and *AMELY* genes to determine genetic sex. Standard PCR amplification of ancestry-specific mtDNA markers (HVRI, HVRII, COII) was performed using a HotStart polymerase and PCR primers that targeted overlapping gene regions. Table 1 is a summary of the mtDNA primers used (with forward and reverse sequences), the optimal annealing temperature for each primer pair, as well as the expected size of the PCR product after amplification. The PCR temperature conditions were as follows: 98 °C, 3 s, (98 °C, 10 s, Tm, 20 s, 72 °C, 15 s) x 46 cycles, 72 °C, 5 min, 10°C, ∞.

**Table 1:** A list of the primers used, target region and expected PCR product size

Primer	Melting temperature (°C)	mtDNA/primer position 3'	Product size (bp)
B1 B1-F CACCATGAATATTGTACGGT B1-R TTGCAGTTGATGTGTGATAG	56	HVRI 16131–16228	140
C1 C1-F AAGTACAGCAATCAACCCTC C1-R CTGTAATGTGCTATGTACGGTA	56	HVRI 16225–16325	141
D3 D3-F TACCCACCCTTAACAGTACA D3-R TATTGATTCACGGAGGA	54	HVRI 16307–16406	136
U2e1 U2e1-F CACAGCCACTTCCACACAG U2e1-R TCTTTGTTTTGGGGTTGG	63	HVRII 274–348	112
Ha Ha-F TCTGAGCCCTAGGATTCATC Ha-R TGATGGCAAATACAGCTCCT	63	COII 6938–7059	153
Hb Hb-F AGACATCGTACTACAGCACAG Hb-R AAGCCTCCTATGATGGCAA	63	COII 7013–7062	90
Amel Amel-F CCCTGGGCTCTGTAAGAATAGTG Amel-R ATCAGAGCTTAACTGGGAAGCTG	66	AMEL	106

PCR was followed by cloning in *Escherichia coli* bacteria, purification and automated Sanger sequencing. CLCBio ([www.clcbio.org](http://www.clcbio.org)) was used to view and edit the sequences and to create contigs. Then the MEGA6 platform<sup>19</sup> was used to create alignments with matching reference sequences from GenBank; the algorithm SplitsTree4 was used to construct an allele network.



## Results

### Radiocarbon dating

Two possible dates were obtained, based on where the samples fell on the calibration curve:  $140 \pm 30$  BP or AD 1675–1735 (2-sigma calibrated) and AD 1800 to post-1950. The earlier date seems more likely based on the few available potsherds, although the cultural material does not help to refine the age of the burial. Further dating, perhaps of the wooden posts, may provide more clarity. Broadly speaking it can be concluded that the age of this burial fell into the Late Iron Age.<sup>20</sup>

### CT scanning

The quality of the CT scans (slice thicknesses of several millimetres) did not allow a definite anatomical-pathological assessment of this individual. Any diagnosis thus has to be taken as provisional.

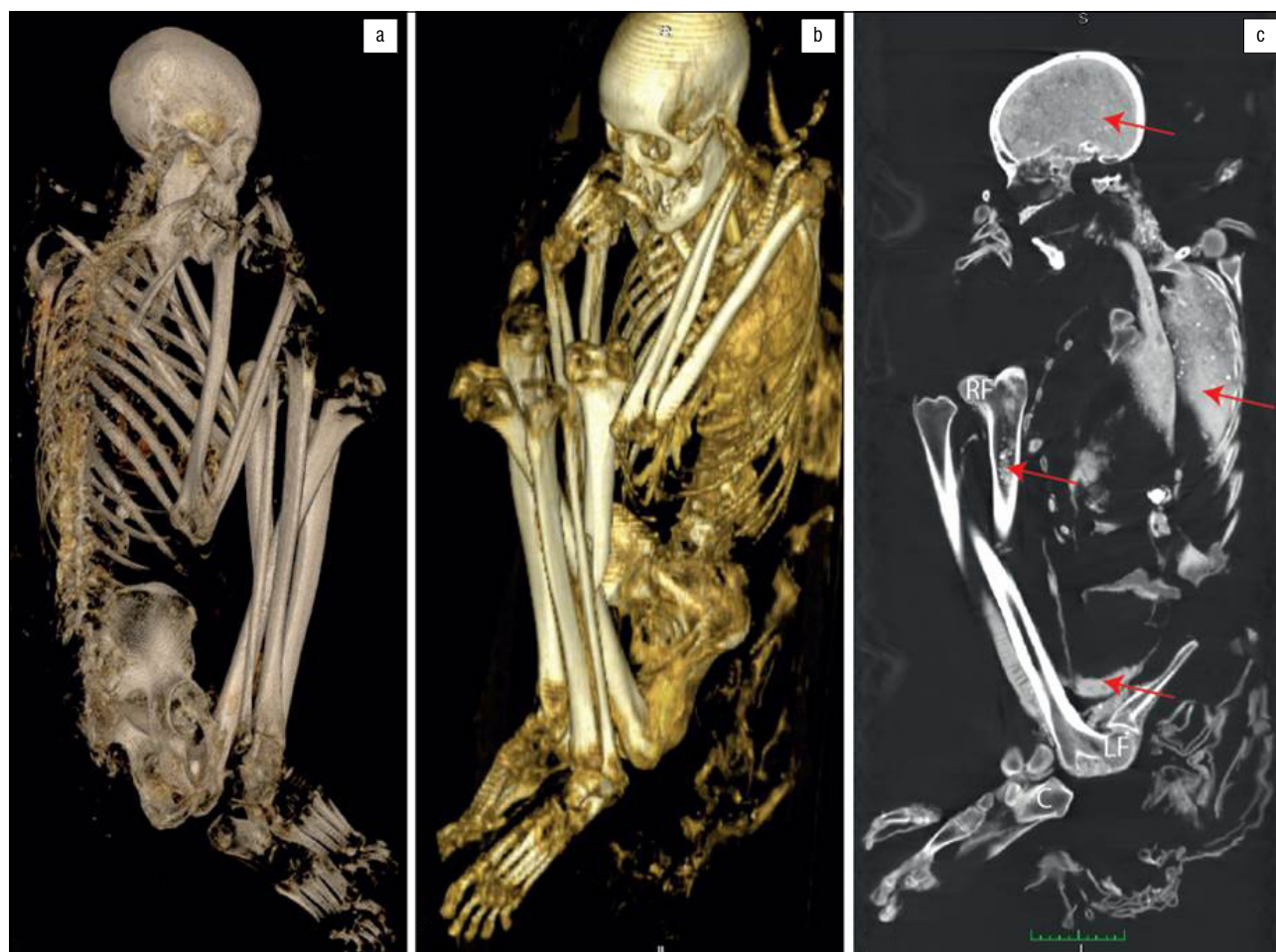
Several bones and bone parts were missing, including the proximal part of the right femur, some carpals, metacarpals and phalanges of the left hand and some phalanges of the right hand. These parts probably were lost as a result of wild animals burrowing into the grave.<sup>9</sup> The mummy was in a foetal position, with the arms and legs flexed and the head inclined forward. The left hand was under the head, close to the angle of the jaw, and the right hand was found in front of the mouth area. The chest and abdomen were heavily compressed, with the entire left side of the chest and abdomen severely damaged. The distal part of the right femur was out of position, in front of the body but rotated. The rotation was a result

of post-mortem damage, most probably caused by animals burrowing into the grave. Figure 2 shows a volume-rendered image of the remains, in which the reversed position of the distal femur can be seen. The rest of the femur had been broken off post-mortem but was recovered next to the burial. Skin lesions and damage were found at multiple parts of the body.

As far as the skull specifically was concerned, the facial bones were found to be intact but no remnants of the brain could be identified (Figure 3). The lamina cribrosa as well as the foramen magnum itself were intact, which mostly excludes artificial excerebration, as seen in, for example, Egyptian mummies. The cranial cavity was almost completely filled with radiodense material (ranging from -177 to +344 HU); only the right cranial cavity showed a 20-mm wide empty region. The infill was probably soil, although some of the debris potentially included brain or dura remnants. The fact that the soil had accumulated on the left side of the cranial cavity with empty space on the right, indicates that the remains had been buried lying on the right side from the start (i.e. it is unlikely that the body fell over after initially having been positioned in a sitting position, for example). The scalp is mostly intact.

All sinuses as well as the mastoids were well pneumatized. The contents of both orbits were not preserved; while the left orbit was empty the right orbit was filled with soil. The mouth was slightly open and empty, but the nasal cavity was filled with soil.

Because of the relatively large slice thicknesses of the CT scans, assessment of the oral structures and dental pathologies was difficult.

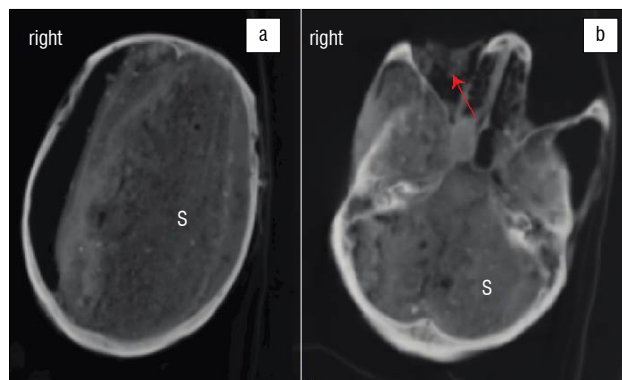


Arrows, soil/stones in cranial and thoracic cavity and pelvis; C, left calcaneal bone; LF, left femoral head; RF, right distal femur in reverse position.

**Figure 2:** (a) Three-dimensional volume rendering, (b) maximal intensity projection and (c) sagittal slice of the Tuli mummy.



Several teeth were lost, probably post-mortem (lower left incisors, upper right incisors, upper left incisors and canine, upper left second premolar, lower left first premolar) (Figure 4). These teeth were not found elsewhere in the wrapping and it is assumed that they may have fallen out and been lost. The alveolar bone of the maxilla and mandible seemed to be lost around the incisival area. It seemed that there was heavy abrasion or wear on the molars, and possibly a deep carious lesion on the occlusal surface of the left lower first molar, with possible ante-mortem loss of the second molar.



S, soil/stones in cranial cavity; arrow, soil/stones in orbital cavity.

Figure 3: (a) Axial slice and (b) coronal slice of the skull of the Tuli mummy.

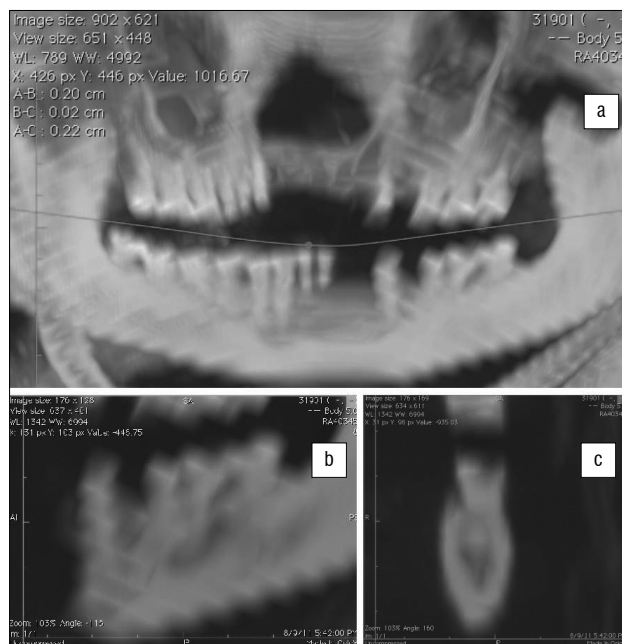


Figure 4: (a) A three-dimensional curved multiplanar reconstruction of the dentition; (b) a two-dimensional multiplanar reconstruction of a sagittal slice of the lower left jaw; and (c) a two-dimensional multiplanar reconstruction of a sagittal slice, with a possible deep carious lesion on the occlusal surface of 36 with apical osteolysis.

The vertebral canal was filled with soil and possibly some remains of dura. The dens of the axis was intact but slightly moved forward within the atlanto-axial joint. The disc spaces from L2 to L5 were widened and an anterolisthesis L4 to L5 (10 mm) was seen (post-mortem alterations). Multiple osteophytes could be observed on the lumbar spine (L2–L5; very pronounced at L3 and L4) (Figure 5), mainly at the right frontal side of the vertebral bodies. No further degenerative changes were seen on the vertebral column.

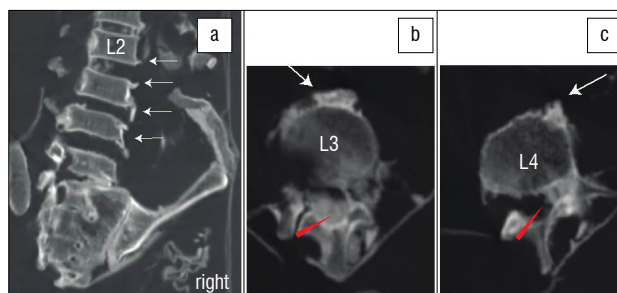


Figure 5: (a) Coronal slice of the lower vertebral column with degenerative alterations (shown by arrows); and axial slices of (b) lumbar vertebra number 3 and (c) lumbar vertebra number 4 (red arrowheads show soil entering the spinal canal).

The pelvic girdle was compressed. While the right hip joint was damaged, the left hip was intact with the head of the femur positioned inside the acetabulum. The joint space was filled with soil. The sacroiliac joints were intact, and some soil entered in the joint space on the left side.

The chest was heavily compressed, mainly on the left side (Figure 6). The left shoulder girdle was moved dorsally whereas the right side was displaced more anteriorly. Both clavicles were rotated dorsally. While the costo-vertebral joints remained intact the ribs were luxated forward in many places and showed multiple fractures. These distortions and fractures are all the result of post-mortem/perimortem damage. The left thoracic cavity was filled with soil; the right thoracic cavity was empty. No internal organs nor the mediastinum could be differentiated. Diverse large skin lesions (most likely post-mortem as a result of handling and taphonomic alterations) were found in the upper chest and neck area.

While the left part of the abdomen was compressed, the right was mainly intact (Figure 6b). No internal organs could be identified. Again, the radiodense material (soil) was seen in the lesser pelvis.



Arrows, dense structures, dry density soil.

Note the absence of internal organs.

Figure 6: Severely compressed (a) thoracic and (b) abdominal cavities.

Both arms were in a flexed position, with the elbow joints intact and showing no degenerative changes. Both knee joints were fully luxated. However, the patello-tibial as well as the proximal tibio-fibular joints remained intact and the right patellar ligament was visible. Both feet were in eversion, with subluxed ankle joints. This position may be the result of tightly flexing the body while wrapping it in the animal skin. Radiodense material was visible in the distal right femur; most likely surrounding soil had entered into the open long bone, but it might also have been liquid material which thickened postmortally.

### Molecular analyses

Some minimal contamination of the PCR reactions was noted, namely by microbial genes such as *Candida*, *Mycobacter* and *Gluconacetobacter*, which was verified through comparison with GenBank sequences. Contaminated samples were not used in further analyses. The DNA sequences obtained were not all usable, and owing to the degraded nature of aDNA, only the HVRI sequences could be used. Sequences were then queried against the GenBank database and apart from the above-mentioned contaminated samples, all shared some sequence similarities with human mtDNA sequences. Among the matches were

mtDNA sequences from a Kgalagadi individual, Sotho, Khoi, Tswana and Zulu individuals. The names of these sequences, named after cultural affinities, were based on self-identification of the donors. Table 2 shows a summary of the alignments between the Tuli mummy and each of the above-mentioned reference sequences. Figure 7 depicts these alignments.

**Table 2:** Summary of variants between query sequence and reference sequences

Reference sequence	Sequence length (bp)	% Identity	Transitions/transversions (R)
Kgalagadi	121	87	0.91
Sotho	121	87	0.91
Tswana	121	86	0.92
Zulu	121	86	0.92

When compared to the revised Cambridge Reference Sequence, the mtDNA sequence obtained from the Tuli mummy was assigned to haplogroup L0, without any further resolution as to which sub-haplogroup it may belong to, owing to the poor quality of the sequences. Therefore, in order to better understand the likely genetic heritage of the Tuli mummy, the sequences were used to create an alignment that also included mtDNA control region sequences from other African sequences, accessed via GenBank.<sup>21</sup> Most of these were southern African and were added for better comparison. This alignment was used to create a

haplotype network, using the program SplitsTree<sup>42</sup>, which is presented in Figure 8. The abbreviations to the sequence labels denote countries of origin or group affinities as reported by anonymous DNA donors.

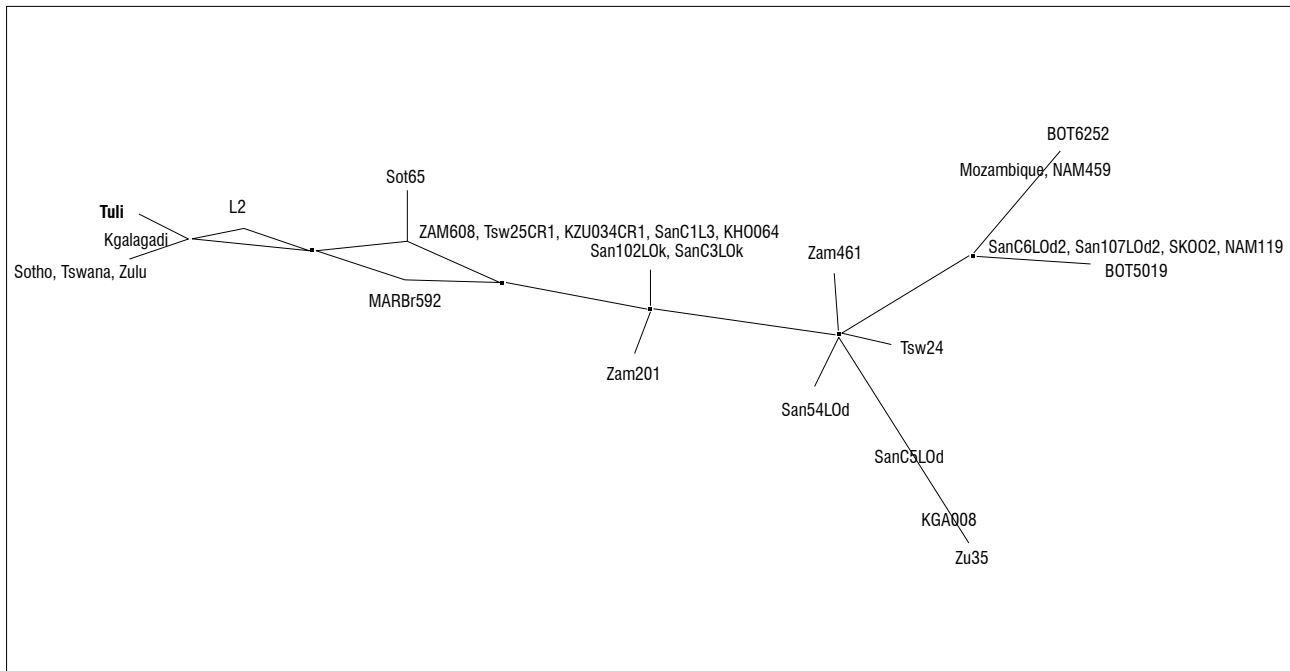
## Discussion

Huffman<sup>20</sup> describes burial practices in Zimbabwe, in which burials of sacred leaders were said to receive special attention. After such a leader died, his body was not immediately interred but instead was mummified, usually by slowly drying it over a low fire. This preserved body would then be wrapped in a cloth or bull hide, and buried at the same time that the successor came into power. It seems possible that a similar process was followed here, and that the body may have been cured or smoked over a low fire which lead to its preservation. What is not clear is whether the internal organs were removed before this process. In the Tuli mummy, no evidence of the presence of internal organs was found. It is possible that the organs have decayed completely, but the possibility that they were removed prior to desiccation should be considered. The extreme flexion of the body is interesting, as is the fact that this level of flexion could be achieved possibly such a long time after death; although wrapping the body in an animal skin and then tying it together with a rope assisted in the process. On the other hand, the fact that other biological material (e.g. buried wooden poles, the skin wrapping the body and plant fibre used to tie the body) was found in good preservation raises the possibility that desiccation of the remains may have occurred naturally.

The individual shows multiple post-mortem alterations, including a compressed thorax/abdomen as well as missing bone parts. However, apart from some degenerative changes of the lower vertebral column, the axial skeleton remained intact. The presence of these multiple degenerative changes warrants a revision of the initial age estimation, indicating that



**Figure 7:** Alignment of the Tuli HVRI sequence against reference sequences representing a (a) Kgalagadi, (b) Sotho, (c) Tswana and (d) Zulu individual, respectively



MARBr592, Morocco; Zu, Zulu; KHO, Khoi; BOT, Botswana; ZAM, Zambia; KGA, Kgalagadi; Tsw, Tswana; NAM, Namibia; L2, haplogroup L2.

**Figure 8:** Network showing the relationship between the Tuli mummy and other reference sequences. The Tuli individual is marked in bold print for easier viewing.

the individual was probably older than what was initially thought. This revision is supported by the observation of advanced dental wear. An age estimate of over 50 years is thus deemed to be more accurate. The Tuli mummy could not be fully evaluated based only on the available CT scans. Difficulties were caused by the slice thickness and partially incongruent scan series, as well as the soil, which entered at diverse regions. Unfortunately, limited facilities are available in Gaborone for this purpose and specialised analyses remain difficult.

The aDNA alignments (Figure 7) show that generally the query sequence, when compared to the reference sequences, presented an 11 nucleotide gap at the query sequence positions 45 to 55, as well as deletions at positions 72 and 105. Only 122 positions could be aligned with relative significance and the transition/transversion bias of 0.91 denotes a higher number of transitions over transversions; however, this ratio is not significant because of the lengths of these alignments and also because of the 11 nucleotide gap.

It is also evident from the median joining network presented in Figure 8 that there is some genetic relatedness between the Tuli mummy and the present day Sotho/Tswana and Khoesan. It also appears that the one Kgalagadi sequence represents an ancestral node that is common to the Tuli sequence as well as that of the Sotho, Tswana and Khoesan. This association is not surprising. The DNA analysis does not imply that the Tuli individual was indeed Khoesan or Sotho/Tswana, but does show at least that there are some mtDNA sequences shared between the Tuli mummy and the present day Khoesan and Sotho/Tswana. It is therefore not impossible that this individual shared maternal genetic heritage with either (or both) of the above-mentioned groups. The similarity shared between this individual and the Zulu is also noteworthy as it could suggest shared ancestry as a result of the migration of different groups (both foragers and farmers), that may have resulted in population interaction and genetic exchange. The assignment of this individual to haplogroup LO only indicates that he was of sub-Saharan African origin, but this is the extent of the secure interpretations because of the poor quality of the sequence obtained. The morphological analyses support this suggestion. Even so, these data will contribute to what will hopefully be a growing database of ancient DNA results from this region – a field which is still in its infancy in southern Africa.

Southern African mummies are rare finds, and detailed analyses even rarer. We have reported the first successful extraction of aDNA from such mummified remains, and included detailed CT scanning results. These may serve as reference for future such studies, which may include other such remains, for example the Kouga mummy.

## Acknowledgements

We acknowledge SANPAD and the Mäxi Foundation (Switzerland) for funding parts of the project. We also thank the radiologists at the Bokamoso Private Hospital for their assistance. The research of M.S. is funded by the National Research Foundation (NRF) of South Africa. Any opinions, findings and conclusions expressed are those of the authors and therefore the NRF does not accept any liability in regard thereto. The permit to analyse the mummy was granted by the Botswana Government through the Botswana National Museum.

## Authors' contributions

F.J.R. is a specialist on mummies and took the samples for aDNA analysis, assisted with the CT scanning and the interpretation thereof, contributed to the writing, and provided funding in Switzerland; M.S. coordinated the project, took the lead in the writing of the manuscript and provided funding in South Africa and Botswana; M.N.M. discovered the mummy, interpreted the dating and archaeology, organised permission in Botswana and coordinated the project in Botswana; L.O. provided expert advice on the CT scans; M.K.B. performed the aDNA analysis (as part of her postgraduate studies) and A.B. oversaw and assisted with the aDNA analysis.

## References

1. Aufderheide AC, Rodríguez-Martín C. The Cambridge encyclopedia of human paleopathology. Cambridge: Cambridge University Press; 1998.
2. Ortner DJ. Identification of pathological conditions in human skeletal remains. 2nd ed. San Diego, CA: Elsevier Academic Press; 2003. <http://dx.doi.org/10.1016/B978-012528628-2/50058-2>
3. Rühli FJ, Chhem RK, Böni T. Diagnostic paleoradiology of mummified tissues: Interpretation and pitfalls. *Can Assoc Radiol J.* 2004;55:218–227.

4. Asingh P, Lynnerup N. Grauballe Man: An Iron Age bog body revisited. Aarhus: Jutland Archaeological Society Publications; 2007. p. 49.
5. Rühli F, Böni T, Perlo J, Casanova F, Baías M, Egarter Vigl E, et al. Non-invasive spatial tissue discrimination in ancient mummies and bones in situ by portable nuclear magnetic resonance. *J Cult Her.* 2007;8:257–263. <http://dx.doi.org/10.1016/j.culher.2007.03.002>
6. Esterhuysen AB, Sanders VM, Smith JM. Human skeletal and mummified remains from AD 1854 siege of Mugobane, Limpopo, South Africa. *J Archaeol Sci.* 2009;36:1038–1049. <http://dx.doi.org/10.1016/j.jas.2008.12.006>
7. Steyn M, Binneman J, Loots M. The Kouga mummified human remains. *S Afr Archaeol Bull.* 2007;62:3–8.
8. Sealy J, Pfeiffer S, Yates R, Willmore K, Manhire A, Maggs T, et al. Hunter-gatherer child burials from the Pakhuis mountains, Western Cape: Growth, diet and burial practices in the Late Holocene. *S Afr Archaeol Bull.* 2000;55:32–43. <http://dx.doi.org/10.2307/3888890>
9. Mosothwane MN. Tuli mummy: A preliminary report from northeastern Botswana. *S Afr Archaeol Bull.* 2011;66:157–160.
10. Lynnerup N. Methods in mummy research. *Anthropol Anz.* 2009;67:357–384. <http://dx.doi.org/10.1127/0003-5548/2009/0028>
11. Knudson KJ, Tung TA, Nystrom KC, Price TD, Fullagar PD. The origin of the Juch'uyupampa Cave mummies: Strontium isotope analysis of archaeological human remains from Bolivia. *J Archaeol Sci.* 2005;32(6):903–913. <http://dx.doi.org/10.1016/j.jas.2005.01.007>
12. Finucane BC. Mummies, maize and manure: Multi-tissue stable isotope analysis of late prehistoric human remains from the Ayacucho Valley, Perú. *J Archaeol Sci.* 2007;34(12):2115–2124. <http://dx.doi.org/10.1016/j.jas.2007.02.006>
13. Handt O, Krings M, Ward RH, Pääbo S. The retrieval of ancient human DNA sequences. *Am J Hum Gen.* 1993;59(2):368–376.
14. Burger J, Hummel S, Hermann B, Henke W. DNA preservation: A microsatellite-DNA study on ancient human skeletal remains. *Electrophoresis.* 1999;20:1722–1728. [http://dx.doi.org/10.1002/\(SICI\)1522-2683\(19990101\)20:8<1722::AID-ELPS1722>3.0.CO;2-4](http://dx.doi.org/10.1002/(SICI)1522-2683(19990101)20:8<1722::AID-ELPS1722>3.0.CO;2-4)
15. Guhl F, Jaramillo C, Vallejo GA, Yockteng R, Cárdenas-Arroyo F, Fornaciari G, et al. Isolation of *Trypanosoma cruzi* DNA in 4,000-year-old mummified human tissue from northern Chile. *Am J Phys Anthropol.* 1999;108:401–407. [http://dx.doi.org/10.1002/\(SICI\)1096-8644\(199904\)108:4<401::AID-AJPA2>3.0.CO;2-P](http://dx.doi.org/10.1002/(SICI)1096-8644(199904)108:4<401::AID-AJPA2>3.0.CO;2-P)
16. Shin DH, Youn M, Chang BS. Histological analysis on the medieval mummy in Korea. *Forensic Sci Int.* 2003;137(2–3):172–182. [http://dx.doi.org/10.1016/S0379-0738\(03\)00335-9](http://dx.doi.org/10.1016/S0379-0738(03)00335-9)
17. Sambrook J, MacCallum P, Russell D. The condensed protocols from molecular cloning: A laboratory manual. New York: Cold Spring Harbor Press; 2006.
18. Suzuki T, Udo K, Adachi M, Sanada H, Tanaka K, Mizuta C, et al. Evolution of the diverse array of phosphagen systems present in annelids. *Comp Biochem Physiol B Biochem Mol Biol.* 2009;152(1):60–66. <http://dx.doi.org/10.1016/j.cbpb.2008.09.087>
19. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol.* 2011;28(10):2731–2739. <http://dx.doi.org/10.1093/molbev/msr121>
20. Huffman TN. Handbook to the Iron Age: The archaeology of pre-colonial farming societies in southern Africa. Pietermaritzburg: University of KwaZulu-Natal Press; 2007.
21. Benson DA, Cavanaugh M, Clark K. GenBank. *Nucleic Acids Res.* 2013;41:36–42. <http://dx.doi.org/10.1093/nar/gks1195>
22. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol.* 2006;23(2):254–267. <http://dx.doi.org/10.1093/molbev/msj030>





# Radioactive nuclides in phosphogypsum from the lowveld region of South Africa

## AUTHORS:

Xolani Msila<sup>1</sup>

Frans Labuschagne<sup>1</sup>

Werner Barnard<sup>2</sup>

David G. Billing<sup>3</sup>

## AFFILIATIONS:

<sup>1</sup>Sasol Dyno Nobel, Bronkhorstspuit, South Africa

<sup>2</sup>Sasol Technology Research and Development, Sasolburg, South Africa

<sup>3</sup>School of Chemistry, University of the Witwatersrand, Johannesburg, South Africa

## CORRESPONDENCE TO:

Xolani Msila

## EMAIL:

xolani.msila@sasol.com

## POSTAL ADDRESS:

Sasol Dyno Nobel, 486  
Brandbach Road, Ekandustria,  
Bronkhorstspuit 1021,  
South Africa

## DATES:

Received: 11 Mar. 2015

Revised: 13 July 2015

Accepted: 01 Aug. 2015

## KEYWORDS:

apatite; radium equivalent;  
radiation hazard indices;  
Anderson–Darling;  
secular equilibrium

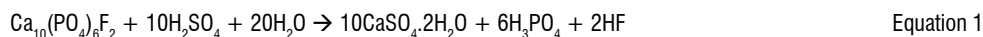
## HOW TO CITE:

Msila X, Labuschagne F, Barnard W, Billing, DG. Radioactive nuclides in phosphogypsum from the lowveld region of South Africa. *S Afr J Sci.* 2016;112(1/2), #Art. 2015-0102, 5 pages. <http://dx.doi.org/10.17159/sajs.2016/20150102>

We evaluated the suitability of phosphogypsum from the Lowveld region of South Africa (LSA), for the manufacturing of building materials, with reference to (1) the *National Nuclear Regulator Act 47 of 1999* and (2) the radioactivity associated risks as quantified in terms of the external and internal hazard indices, the activity concentration index and the radium equivalent. The distribution of radioactive nuclides in the LSA phosphogypsum was also examined. Analyses of 19 samples of the phosphogypsum show that phosphogypsum contains lower activity concentrations of naturally occurring radioactive nuclides of uranium and thorium and their progeny than the 500 Bq/kg limit set for regulation in South Africa. The potassium-40 (<sup>40</sup>K) activity concentration was below the minimum detectable amount of 100 Bq/kg. The values obtained for external and internal hazard indices and the activity concentration index were:  $2.12 \pm 0.59$ ,  $3.44 \pm 0.64$  and  $2.65 \pm 0.76$  respectively. The calculated radium equivalent  $Ra_{eq}$  was  $513 \pm 76$  Bq/kg. The final decision regarding phosphogypsum's suitability for use as a building material should consider scenarios of use.

## Introduction

A phosphoric acid production facility has been in operation in the Lowveld region of South Africa (LSA) since the 1960s. The process of producing phosphoric acid involves the digestion of fluoro-apatite ore ( $Ca_{10}(PO_4)_6F_2$ ) with sulfuric acid as shown in Equation 1:



Each year, tons of calcium sulfate dihydrate ( $CaSO_4 \cdot 2H_2O$ ) or gypsum, or specifically phosphogypsum, is produced as a by-product. The LSA phosphogypsum is stored in waste stacks alongside the phosphoric acid factory. In a country like South Africa, that has a challenge to provide low cost or affordable housing,<sup>1</sup> the use of phosphogypsum to manufacture building material is an attractive option. The manufacturing of low cost prefabricated building material from gypsum has in the recent past been demonstrated by Rajkovic and Toskovic<sup>2</sup>. But the solubility of uranium, thorium and their daughter products that exist in apatite ores<sup>3,4</sup> result in these radioactive nuclides partitioning between the phosphoric acid and the phosphogypsum<sup>5,6</sup> during the processing of the ore. Rajkovic and Toskovic<sup>2</sup>, Al-Jundi et al.<sup>6</sup> and Hussein<sup>7</sup> studied phosphogypsum from Serbia, Jordan and Egypt respectively. Hussein did not report on the activity concentrations of potassium-40 (<sup>40</sup>K), thorium-232 (<sup>232</sup>Th) and radium-226 (<sup>226</sup>Ra) but the results of radioactive nuclides activities from Serbia and Jordan are given in Table 1.

**Table 1:** Activity concentrations (Bq/kg) of <sup>40</sup>K, <sup>232</sup>Th and <sup>226</sup>Ra in phosphogypsum from other countries

Country	Number of samples	<sup>40</sup> K	<sup>232</sup> Th	<sup>226</sup> Ra
Serbia <sup>2</sup>	-	8.7	8.7	439
Jordan <sup>6</sup>	15	40	-	376

The worldwide average activities of <sup>226</sup>Ra, <sup>232</sup>Th and <sup>40</sup>K on the earth's crust are about: 40; 40 and 400 Bq/kg, respectively.<sup>8</sup> Humans are therefore exposed to this naturally occurring radioactive material (NORM). If the material used for the construction of human dwellings contributes additional radioactivity to the naturally occurring radioactive material, radiation exposure increases. The presence of these radioactive nuclides impurities in the LSA phosphogypsum can limit its use as a building material.

In this study we evaluated the suitability of the LSA phosphogypsum for the manufacturing of building materials, with reference to (1) the *National Nuclear Regulator Act 47 of 1999*<sup>9</sup> and (2) the radioactivity associated risks as quantified in terms of the external ( $H_{ex}$ ) and internal ( $H_i$ ) hazard indices, the activity concentration index ( $I_a$ ) and the radium equivalent ( $Ra_{eq}$ ). The effect of ore processing on the distribution of the radioactive nuclides was also examined.

## Materials and method

During the production of phosphoric acid for analysis, 19 grab samples of LSA phosphogypsum were taken.

### Sample preparation

The phosphogypsum was dried overnight in an oven at 80 °C. About 100 g of the dried phosphogypsum was transferred to a grinding vessel (containing stainless steel grinding balls and block) with a swing mill. Milling was performed at 960 rpm. for one min to obtain a fine powder. The process was repeated until 0.5 kg of homogeneous fine powder was generated.

### Gamma analysis

A standard glass container was filled with phosphogypsum and the mass recorded. The container was closed with a lid and sealed airtight with the aid of epoxy resin. The prepared sample was allowed to stand for 3 weeks so that radon-222 (and some of its progeny) reached radioactive equilibrium with radium-226. The  $\gamma$ -spectrum was acquired on a high purity germanium detector calibrated for the standard glass container geometry. The detector was housed in a lead shield to reduce background radiation. Gamma spectrum analysis was performed with Genie™ 2000 Model S501 Gamma Analysis software. The weighted average activity of bismuth-214 and lead-214 was calculated and reported for radium. Thorium-228 was calculated from lead-212 and thallium-208. Radium-228 was determined from actinium-228. Potassium-40 was measured directly.

### Uranium and thorium analysis by activation analysis

Aliquots of the powder were transferred into irradiation capsules. Uranium and thorium standards were prepared by transferring known amounts of these elements from certified reference solutions into capsules and evaporated to dryness. Samples and standards were sequentially transferred with a pneumatic system into an in-core position of the SAFARI-1 reactor and irradiated with neutrons for a fixed time. After irradiation, the sample was transferred to a detector to measure neutrons emitted by products formed from fission of uranium-235.

The emission of neutrons between samples and standards was compared to calculate the uranium concentration in the sample. After a prolonged period, the irradiated samples and standard were measured on a gamma detector to determine neptunium-239 and protactinium-231. These nuclides formed from the neutron activation of uranium-238 and thorium-232 respectively. By comparing the activities of the nuclides in the standards and samples, the uranium-238 and thorium-232 sample activities were calculated.

### Radiation indices

From the activity concentrations of the radioactive nuclides,  $H_{ex}$ ,  $H_i$ ,  $I_y$  and  $Ra_{eq}$  were calculated using the formulae in Equations 2–5 respectively. The formulae and their applications are comprehensively defined by the European Commission for Radiation Protection<sup>8</sup>.

$$H_{ex} = \frac{C_{Ra}}{370} + \frac{C_{Th}}{258} + \frac{C_K}{4180} \quad \text{Equation 2}$$

$$H_{in} = \frac{C_{Ra}}{185} + \frac{C_{Th}}{259} + \frac{C_K}{4810} \quad \text{Equation 3}$$

$$I_y = \frac{C_{Ra}}{300} + \frac{C_{Th}}{200} + \frac{C_K}{3000} \quad \text{Equation 4}$$

$$Ra_{eq} = C_{Ra} + 1.43C_{Th} + 0.077C_K \quad \text{Equation 5}$$

where  $C_{Ra}$ ,  $C_{Th}$  and  $C_K$  are activities of <sup>226</sup>Ra, <sup>232</sup>Th and <sup>40</sup>K in Bq/kg, respectively.<sup>10,11</sup>

## Results and discussion

### Radionuclide concentrations in phosphogypsum

The average activity concentrations of the radioactive nuclides: uranium-238 (<sup>238</sup>U), lead-210 (<sup>210</sup>Pb), uranium-235 (<sup>235</sup>U), radium-228 (<sup>228</sup>Ra), thorium-232 (<sup>232</sup>Th), <sup>226</sup>Ra, <sup>232</sup>Th and <sup>40</sup>K in the LSA phosphogypsum are presented in Table 2. According to the *National Nuclear Regulator Act, 47 of 1999*,<sup>9</sup> an operation or material is excluded from regulation if activity concentration of the naturally occurring radioactive nuclides of uranium, thorium and their progeny are each below 500 Bq/kg. The limit for <sup>40</sup>K is 10 000 Bq/kg.

The results confirm that the activity concentrations measured in the 19 LSA phosphogypsum samples are all below the regulation limits set by the *National Nuclear Regulator Act 47 of 1999*<sup>9</sup> for <sup>238</sup>U, <sup>235</sup>U, <sup>210</sup>Pb, <sup>228</sup>Ra, <sup>226</sup>Ra, <sup>228</sup>Th, and <sup>232</sup>Th. The activity concentration of <sup>40</sup>K in the LSA phosphogypsum was determined to be below the minimum detectable limit of 100 Bq/kg. Numerous studies have revealed that building materials contain appreciable activity concentrations of radioactive nuclides <sup>226</sup>Ra, <sup>232</sup>Th and <sup>40</sup>K. There are limited data available on the radiological safety of phosphogypsum use as a building material beyond the study by Rajkovic and Toskovic<sup>2</sup> and therefore to put the results obtained for LSA phosphogypsum into perspective, some of the results obtained for bricks from several countries are presented in Table 3.

**Table 3:** Activity concentrations (Bq/kg) of <sup>40</sup>K, <sup>232</sup>Th and <sup>226</sup>Ra in building bricks from various countries

Country	Number of samples	<sup>226</sup> Ra	<sup>232</sup> Th	<sup>40</sup> K
Norway <sup>12</sup>	6	104	62	1058
Greece <sup>13</sup>	6	49	24	670
Australia <sup>14</sup>	25	41	89	681
Netherlands <sup>15</sup>	14	39	41	560
Egypt <sup>16</sup>	1	20	14	204
India <sup>17</sup>	1	48	52	381
Sri Lanka <sup>18</sup>	24	35	72	585
India <sup>19</sup>	Not reported	18	19	238
India <sup>20</sup>	Not reported	47	20	349

The average activity concentration of <sup>226</sup>Ra in the bricks listed in Table 3 range from 18 Bq/kg to 104 Bq/kg. This is lower than the average activity concentration in the LSA phosphogypsum (109 Bq/kg ± 18 Bq/kg). Similarly, the average activity concentration of <sup>232</sup>Th in the bricks ranges from 14 Bq/kg to 89 Bq/kg which is lower than the average in the LSA phosphogypsum, where <sup>232</sup>Th is 253 Bq/kg ± 160 Bq/kg. The activity concentration of <sup>40</sup>K (<100 Bq/kg) is lower in the LSA phosphogypsum

**Table 2:** Activity concentrations of radioactive nuclides in the phosphogypsum from the Lowveld region of South Africa

Radioactive nuclide	<sup>238</sup> U	<sup>210</sup> Pb	<sup>235</sup> U	<sup>228</sup> Ra	<sup>228</sup> Th	<sup>226</sup> Ra	<sup>232</sup> Th	<sup>40</sup> K
	Bq/kg							
Average	31	135	1	404	253	109	189	<100
Standard deviation	20	37	1	70	160	18	155	-
Number of samples	19							

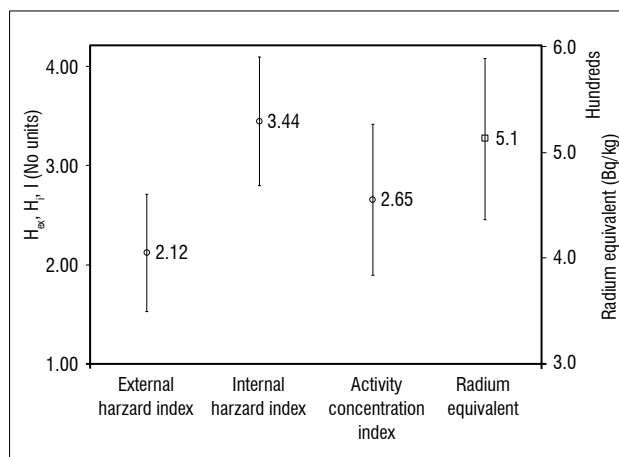
compared to the range of activity concentrations reported for the bricks from other countries (204–1058 Bq/kg).

**Table 4:** The scenarios of use for building materials at different dose criteria

Dose criteria	0.3 mS/year	1 mS/year
Material used in bulk amounts, e.g. bricks	$I_v \leq 0.5$	$I_v \leq 1$
Superficial and other material with restricted use: tiles, and boards	$I_v \leq 2$	$I_v \leq 6$

### Radiation hazard indices

The LSA phosphogypsum is rich in  $^{228}\text{Ra}$  compared to  $^{232}\text{Th}$  and as a result the activity concentration of the former is used in the place of the activity of the latter to calculate the hazard indices in Equations 2–4. The calculated hazard indices  $H_{\text{ex}}$ ,  $H_{\text{in}}$ ,  $I_v$  and the  $\text{Ra}_{\text{eq}}$  are plotted in Figure 1. The error bars in Figure 1 represent the standard deviation of the results of the 19 LSA phosphogypsum samples. The values obtained for  $H_{\text{ex}}$ ,  $H_{\text{in}}$  and  $I_v$  are  $2.12 \pm 0.59$ ,  $3.44 \pm 0.64$  and  $2.65 \pm 0.76$  respectively. The calculated  $\text{Ra}_{\text{eq}}$  is  $513 \pm 76$  Bq/kg.



**Figure 1:** The calculated external ( $H_{\text{ex}}$ ) and internal ( $H_{\text{in}}$ ) hazard indices, activity concentration index ( $I_v$ ) and the radium equivalent ( $\text{Ra}_{\text{eq}}$ ) in phosphogypsum from the Lowveld region of South Africa. Error lines indicate the standard deviation.

The assessment of a material's suitability for use as a building material should be based on scenarios where the material is used. The scenarios for use at different dose criteria<sup>8</sup> are given in Table 4. The activity concentration index,  $I_v$ , should be evaluated against these criteria. If the  $I_v$  is 1 or less, the material can be used as building material, without restriction, as far as radioactivity is concerned, whereas if the  $I_v$  is above 1 and less or equal to 6, the material should be used superficially.

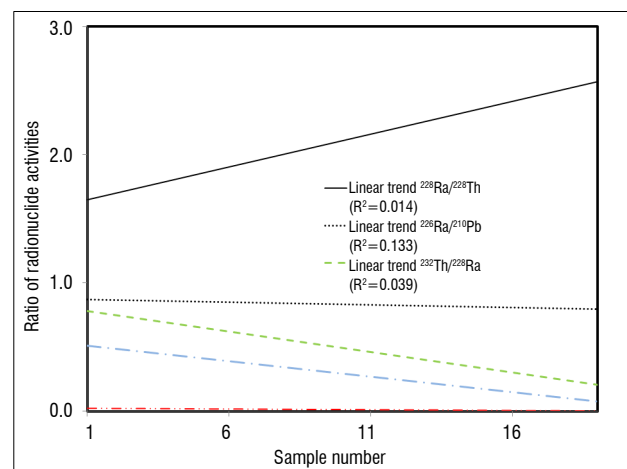
The  $\text{Ra}_{\text{eq}}$  as defined in Equation 5, is calculated from the activity concentrations of  $^{228}\text{Ra}$ ,  $^{232}\text{Th}$  and  $^{40}\text{K}$ . Equation 2 is based on the estimation that 1 Bq/kg of  $^{226}\text{Ra}$ , 0.7 Bq/kg of  $^{232}\text{Th}$  and 13 Bq/kg of  $^{40}\text{K}$  generate the same  $\gamma$ -rays dose rate.<sup>10,21,22</sup> According to Mondal et al.<sup>23</sup> and supported by El-Taher and Makluf<sup>24</sup>,  $I_v=1$  is equivalent to the  $\text{Ra}_{\text{eq}}$  of 370 Bq/kg. The calculated  $\text{Ra}_{\text{eq}}$  of  $513 \pm 76$  Bq/kg for LSA phosphogypsum is consistent with the activity concentration ( $1 < I_v < 6$ ).

### Radioactive nuclides distribution in LSA phosphogypsum

A simplified definition of secular equilibrium is found in Zhang et al.<sup>25</sup> Activity of daughter radionuclides build up to that of the parent in about seven half-lives and thereafter, parent activity ( $A_1$ ) is the same as the activity of its progeny ( $A_2$ ). It follows therefore that at the state of secular equilibrium, the ratio of the activity of the parent to that of the daughter is 1 ( $A_1/A_2 = 1$ ). The secular equilibrium observed in natural ores can be disturbed by mineral processing. Al-Jundi et al.<sup>7</sup> showed that concentration of  $^{238}\text{U}$  and its decay products  $^{210}\text{Pb}$  and  $^{226}\text{Ra}$  originating

from apatite ore are partitioned during processing in such a way that  $^{238}\text{U}$  accumulates in the phosphate fertiliser while  $^{228}\text{Ra}$ ,  $^{226}\text{Ra}$  and  $^{210}\text{Pb}$  accumulate in the phosphogypsum. This partitioning behaviour can be attributed to the very low aqueous solubility of radium sulfate and lead sulfate relative to that of uranium sulfate as almost all the  $^{228}\text{Ra}$  and  $^{226}\text{Ra}$  reported by Van der Westhuizen<sup>26</sup> as present in the phosphate rock is observed in the LSA phosphogypsum.

The ratios of activity concentrations in the LSA phosphogypsum for  $^{238}\text{U}$ ,  $^{235}\text{U}$ ,  $^{232}\text{Th}$  and their progeny in the natural decay series are plotted in Figure 2. The  $^{226}\text{Ra}$  and  $^{210}\text{Pb}$  belong to the  $^{238}\text{U}$  natural decay series and are not in secular equilibrium in the LSA phosphogypsum. Although the linear trend of the ratio of these two radionuclides is close to unity, there is a poor correlation between their activities as shown by the regression coefficient ( $R^2$ ) of 0.133. The uranium radioactive nuclide  $^{238}\text{U}$  is also in disequilibrium with its progeny  $^{226}\text{Ra}$ . Similarly,  $^{235}\text{U}$  is in disequilibrium with  $^{210}\text{Pb}$  and disequilibrium is also observed between  $^{232}\text{Th}$  versus  $^{228}\text{Ra}$  and  $^{228}\text{Ra}$  versus  $^{228}\text{Th}$ .



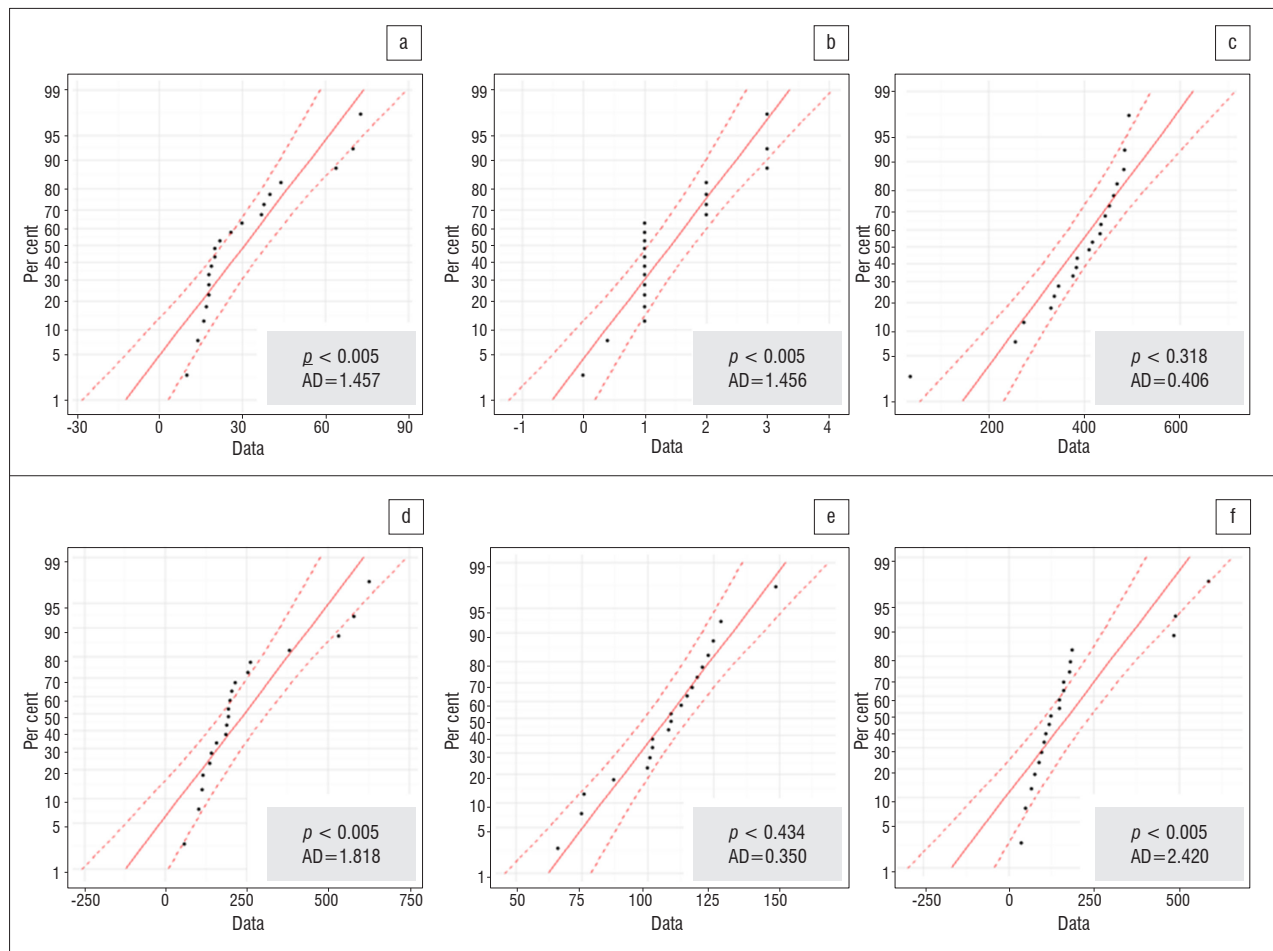
**Figure 2.** The ratio of activity concentrations for  $^{238}\text{U}$ ,  $^{235}\text{U}$ ,  $^{232}\text{Th}$  and their progeny in phosphogypsum from the Lowveld region of South Africa.

The Anderson–Darling test has been shown to give appropriate empirical distribution function statistics for detecting departure from normality, even with small samples ( $n \leq 25$ ).<sup>27</sup> The test was performed on a set of data to evaluate their fit to a chosen distribution pattern. For the Gaussian distribution test, the statistics are based on the squared difference between the normal and the empirical data. If the calculated  $p$ -value is less than a chosen alpha (one minus the confidence interval), the null hypothesis, that the data come from that distribution, is rejected. The Anderson–Darling (AD) test was performed on the radioactive nuclide activity concentration data using the Minitab<sup>®</sup> 16 statistical software to evaluate the distribution pattern of the activity concentration in the LSA phosphogypsum. The results obtained are presented in the probability plot in Figure 3.

At 95% confidence level, a  $p$  value of less than 0.05 indicates no deviation from the Gaussian distribution. This behaviour (normal distribution) is observed for the activity concentration of the radioactive nuclides  $^{238}\text{U}$ ,  $^{235}\text{U}$ ,  $^{228}\text{Th}$  and  $^{232}\text{Th}$  in the LSA phosphogypsum. The contrary is observed for the  $^{228}\text{Ra}$  and  $^{226}\text{Ra}$  activity concentrations. The calculated  $p$  values are 0.318 and 0.434 respectively and the Anderson–Darling values are 0.406 and 0.350 respectively. This can be attributed to the low solubility of radium sulfate which results in accumulation in the phosphogypsum rather than the phosphoric acid.

### Conclusion and recommendation

The LSA phosphogypsum contains appreciable amounts of the radioactive nuclides  $^{238}\text{U}$ ,  $^{210}\text{Pb}$ ,  $^{235}\text{U}$ ,  $^{228}\text{Ra}$ ,  $^{228}\text{Th}$ ,  $^{226}\text{Ra}$ ,  $^{232}\text{Th}$  and below-detectable levels of  $^{40}\text{K}$ . The average activities are below the limits set by the *National Nuclear Regulator Act 47 of 1999*.<sup>9</sup> The phosphogypsum is therefore excluded from regulatory control. The calculated results of



**Figure 3:** Probability distribution for the activity concentration of (a)  $^{238}\text{U}$ , (b)  $^{235}\text{U}$ , (c)  $^{228}\text{Ra}$ , (d)  $^{228}\text{Th}$ , (e)  $^{226}\text{Ra}$  and (f)  $^{232}\text{Th}$ . In all figures, the data on x-axis are the activity concentrations in Bq/kg. The Anderson–Darling (AD) statistic for each radioactive nuclide is displayed on each figure.

hazard indices:  $H_{\text{ex}}$ ,  $H_{\text{in}}$ ,  $I_{\text{v}}$  and  $Ra_{\text{eq}}$  indicate that the LSA phosphogypsum can be utilised as building material if used superficially or with restriction. A final decision on the usability of the material can be made when the scenario of use is known and a more representative sample of the aged bulk material is analysed. The radioactive nuclides and their progenies are not in secular equilibrium in the LSA phosphogypsum and the distribution  $^{238}\text{U}$ ,  $^{235}\text{U}$ ,  $^{232}\text{Th}$  and  $^{228}\text{Th}$  deviates from Gaussian.

## Acknowledgements

Many thanks to Deon Kotze (NECSA) for performing some of the radioactive nuclides analyses.

## Authors' contributions

D.G.B. and W.B. were the project leaders, F.L. was responsible for sampling and arranging for analysis. X.M. handled the arrangement for sample analysis, performed the calculations and wrote the manuscript.

## References

- Ajayi JR. Strategies for sustainable housing cooperatives in South Africa. [PhD thesis]. Port Elizabeth: Nelson Mandela Metropolitan University; 2012.
- Rajkovic MB, Toskovic DV. Investigation of the possibilities of phosphogypsum application for building partitioning walls – elements of a prefabricated house. *APTEFF*. 2002;33:71–92. <http://dx.doi.org/10.2298/APT0233071R>
- Pereira F, Bilal E. Phosphoric acid extraction and rare earth recovery from apatites of the Brazilian phosphate ores. *Romanian J Miner Dep*. 2012;85(2):49–52.
- Garcia-Talavera M, Leadermann JP, Decombaz M, Daza MJ, Quintana B. Coincidence summing corrections for the natural decay series in  $\gamma$ -ray spectrometry. *J Radiat Isotopes*. 2001;54:769–776. [http://dx.doi.org/10.1016/S0969-8043\(00\)00318-3](http://dx.doi.org/10.1016/S0969-8043(00)00318-3)
- Dias NMP, Caires EF, Pires LF, Bacchi MA, Fernandes EAN. Radiological impact of phosphogypsum surface application in a no-till system in southern Brazil. *Pesq Agropec bras Brasilia*. 2010;45:1456–1464. <http://dx.doi.org/10.1590/S0100-204X2010001200017>
- Al-Jundi J, Al-Ahmad N, Shehadeh H, Afaneh F, Maghrabi M, Gerstmann U. Investigation on the activity concentrations of U-238, Ra-226, Pb-210 and K-40 in Jordan phosphogypsum and fertilizer. *Radiat Prot Dosim*. 2008;131(4):449–454. <http://dx.doi.org/10.1093/rpd/ncn214>
- Hussein EM. Radioactivity of phosphate ore, superphosphate and phosphogypsum in Abu Zaabal phosphate plant, Egypt. *Health Phys*. 1994;48:87–95. <http://dx.doi.org/10.1097/00004032-199409000-00010>
- European Commission. Radiation Protection 112: Radiological protection principles concerning the natural radioactivity of building materials. Brussels: Directorate-General: Environment, nuclear safety and civil production; 1999.
- Department of Energy. National Nuclear Regulator Act 47 of 1999 [document on the Internet]. c1999 [cited 2015 December 04]. Available from: <http://www.energy.gov.za/files/esources/nuclear/act47.pdf>
- Mehdizadeh S, Faghihi R, Sina S. Natural radioactivity in building materials in Iran. *Nukleonika*. 2011;56(4):363–368.
- Radenkovic MB, Alshikh SM, Andric VB, Miljanic SS. Radioactivity of sand from several renowned public beaches and assessment of the corresponding environmental risks. *J. Serb Chem Soc*. 2009;74(4):461–470. <http://dx.doi.org/10.2298/JSC0904461R>



12. Papastefanou C, Manolopolou M, Charalambous S. Exposure from the radioactivity in the building materials. *Health Phys.* 1983;45:349–361.
13. Beretka J, Mathew PJ. Natural radioactivity of Australian building materials, industrial waste by-products. *Health Phys.* 1985;48(1):87–95. <http://dx.doi.org/10.1097/00004032-198501000-00007>
14. Ackers JG, Den Boer JF, De Jong P, Wolschrijn RA. Radioactivity and radon exhalation rates of building materials in the Netherlands. *Sci Total Environ.* 1985;45:151–156. [http://dx.doi.org/10.1016/0048-9697\(85\)90215-3](http://dx.doi.org/10.1016/0048-9697(85)90215-3)
15. El-Thaway MS, Higgy RH. Natural radioactivity in different types of bricks fabricated and used in the Cairo region. *Appl Radiat Isotopes.* 1985;46(12):1401–1406. [http://dx.doi.org/10.1016/0969-8043\(95\)00220-8](http://dx.doi.org/10.1016/0969-8043(95)00220-8)
16. Kumar V, Ramachandra TV, Prasad R. Natural radioactivity of Indian building materials and by products. *Appl Radiat Isotopes.* 1999;51(1):93–96. [http://dx.doi.org/10.1016/S0969-8043\(98\)00154-7](http://dx.doi.org/10.1016/S0969-8043(98)00154-7)
17. Hewamanna R, Sumithrararachchi CS, Mahawatte P, Nanayakkara HLC, Ratnayakhe HC. Natural radioactivity and gamma dose from Sri Lankan clay bricks used in building construction. *Appl Radiat Isotopes.* 2001;54:365–369. [http://dx.doi.org/10.1016/S0969-8043\(00\)00107-X](http://dx.doi.org/10.1016/S0969-8043(00)00107-X)
18. Ravisankar R, Vanasundari K, Chandrasekaran A, Suganya M, Eswaran P, Vijayagopal P. Measurement of natural radioactivity in brick samples of Namakkal, Tamilnadu, India using gamma-ray spectrometry. *Arch Phys Res.* 2011;2(2):95–99.
19. Viruthagiri G, Ponnarasi K. Measurement of natural radioactivity in bricks samples. *Adv Appl Sci Res.* 2011;2(2):103–108.
20. Dhanya B, Umadevi AG, Jose PA, Rajagopala M, Jojo PJ. a study on activity concentration of natural radionuclide of building materials in Kochi. *Int J Fund Phy Sci.* 2012;2(3):41–43.
21. Uosif MAM. Specific activity of  $^{226}\text{Ra}$ ,  $^{232}\text{Th}$ , and  $^{40}\text{K}$  for assessment of radiation hazards from building materials commonly used in Upper Egypt. *SDU J Sci.* 2011;6(2):120–126.
22. Le Nhu S, Nguyen TB, Truong Y, Nguyen TN, Nguyen TL, Nguyen VP. Natural radioactivity in common building materials used in Vietnam 11255. *WM2011 Conference*; 2011 February 27–March 03; Phoenix, AZ; 2011.
23. Mondal T, Sengupta D, Mandal A. Natural radioactivity of ash and coal in major thermal power plants of West Bengal, India. *Curr Sci.* 2006;91(10):1387–1393
24. El-Taher A, Makhluf S. Natural radioactivity levels in phosphate fertilizer and its environmental implications in Assuit governorate, Upper Egypt. *Indian J Pure Ap Phy.* 2010;48:697–702.
25. Zhang Q, Burman C, Amols H. Transient and secular radioactive equilibrium revisited [article on the Internet]. c2014 [cited 2015 December 04]. Available from:<http://arxiv.org/ftp/arxiv/papers> (accessed: 4 December 2015)
26. Van der Westhuizen AJ. Rail transport of igneous phosphate rock. In: *International Atomic Energy Agency (IAEA) conference on naturally occurring radioactive materials (NORM IV)*; 2004 May 17-21; Szczyrk, Poland. Vienna: IAEA; 2004.
27. Stephens MA. EDF Statistics for goodness of fit and some comparisons. *J Am Stat Assoc.* 1974;69:730–737 <http://dx.doi.org/10.1080/01621459.1974.10480196>



# Investigating atmospheric photochemistry in the Johannesburg-Pretoria megacity using a box model

## AUTHORS:

Alexandra S.M. Lourens<sup>1</sup>

Tim M. Butler<sup>2</sup>

J. Paul Beukes<sup>1</sup>

Pieter G. van Zyl<sup>3</sup>

Gerhard D. Fourie<sup>4</sup>

Mark G. Lawrence<sup>2</sup>

## AFFILIATIONS:

<sup>1</sup>Unit for Environmental Sciences and Management, North-West University, Potchefstroom, South Africa

<sup>2</sup>Institute for Advanced Sustainability Studies, Potsdam, Germany

<sup>3</sup>School of Physical and Chemical Sciences, North-West University, Potchefstroom, South Africa

<sup>4</sup>EnviroNgaka, Brits, South Africa

## CORRESPONDENCE TO:

Pieter van Zyl

## EMAIL:

pieter.vanzyl@nwu.ac.za

## POSTAL ADDRESS:

School of Physical and Chemical Sciences, North-West University, Hoffman Street Potchefstroom 2520, South Africa

## DATES:

Received: 06 May 2015

Revised: 31 July 2015

Accepted: 01 Aug. 2015

## KEYWORDS:

megacities; air pollution; photochemical box model; ozone; NOx

## HOW TO CITE:

Lourens ASM, Butler TM, Beukes JP, Van Zyl PG, Fourie GD, Lawrence MG. Investigating atmospheric photochemistry in the Johannesburg-Pretoria megacity using a box model. *S Afr J Sci.* 2016;112(1/2), Art. #2015-0169, 11 pages. <http://dx.doi.org/10.17159/sajs.2016/2015-0169>

© 2016. The Author(s).  
Published under a Creative Commons Attribution Licence.

Urban air pollution has become a major concern over the past decades. One of the largest conurbations in Sub-Saharan Africa is developed around the cities of Johannesburg and Pretoria (Jhb-Pta megacity). In this study, a photochemical box model with a detailed representation of ozone (O<sub>3</sub>) formation chemistry was used to investigate the state of current air quality and photochemical processes in the Jhb-Pta megacity, as well as scenarios that could possibly mitigate air pollution. Results indicated that the Jhb-Pta megacity is within a VOC-limited (or NOx-saturated) regime. Major sources of NOx include transport from the Mpumalanga Highveld and local traffic emissions. O<sub>3</sub> levels in the Jhb-Pta megacity will be more effectively reduced if VOC (volatile organic compound) emissions are decreased. A reduction of NOx emissions leads to an increase in O<sub>3</sub> because of a decrease in titration through the reaction with NO. The same effect was observed in various cities worldwide where O<sub>3</sub> levels increased when NOx emissions were reduced during emission control strategies. The effect of reducing vehicular emissions in the Jhb-Pta megacity on the production of O<sub>3</sub> was also investigated. A significant increase of approximately 23 ppb O<sub>3</sub> was observed when emissions of VOCs, NOx and CO were reduced by changing from Euro-0 to Euro-3 vehicles. It is therefore recommended that VOC emissions are decreased together with the implementation of Euro-3 and cleaner vehicles in the future.

## Introduction

One of the largest conurbations in Sub-Saharan Africa is developed around the cities of Johannesburg and Pretoria. This conurbation is one of the 40 largest metropolitan areas in the world with a population of over 10 million. Conurbations of this magnitude have been defined as megacities.<sup>1</sup> The Johannesburg-Pretoria (Jhb-Pta) megacity is also the central hub of economic activity in southern Africa. The main pollution sources in the Jhb-Pta megacity have been identified as traffic emissions, biomass and domestic combustion (space heating and cooking), as well as industrial activities.<sup>2,3</sup> The major pollutants emitted from these activities include nitrogen oxide (NO), nitrogen dioxide (NO<sub>2</sub>), sulphur dioxide (SO<sub>2</sub>), carbon monoxide (CO), particulate matter (PM) and various organic compounds. The Jhb-Pta megacity is also located relatively close to large industrialised regions in South Africa, i.e. the Mpumalanga Highveld, approximately 100 km to the west and the Vaal Triangle approximately 50 km to the south. It is therefore likely that air quality in the megacity will be influenced by the proximity of these areas.

According to the National Environmental Management Air Quality Act<sup>4</sup>, it is a requirement to utilise air quality models in decision-making processes and impact studies in South Africa. Implementation of this requirement has thus far mainly focused on applying dispersion modelling and not on improving processes within these models. However, these models are seldom developed and modified for the unique South African conditions, which differ vastly from conditions in America and/or Europe where most of these models were developed. Some of the dispersion models also include very limited chemical mechanisms.

Photochemical box models with a comprehensive chemical scheme are very useful tools to investigate air quality and the chemical processes within a region and are used to explore the main reactions that participate in O<sub>3</sub> (ozone) formation, as well as to assist in projecting the success of legislative air quality goals resulting from the implementation of emission control plans.<sup>5</sup> Since the first urban scale photochemical box model was successfully developed and used to assess air quality in urban areas,<sup>6</sup> various other photochemical models have been developed to study the production of O<sub>3</sub> in urban and rural areas.<sup>7-9</sup> Young and co-workers<sup>10</sup>, for example, developed and used a photochemical box model to investigate the best emission control strategy to reduce O<sub>3</sub> levels in Mexico City. Box models have also been applied in larger geographic regions.<sup>11-14</sup>

In an effort to investigate the interaction of O<sub>3</sub> precursor species and the resulting O<sub>3</sub> levels in the Jhb-Pta megacity, as well as to determine the possible contribution of O<sub>3</sub> precursor species emissions in areas surrounding the Jhb-Pta megacity, an existing photochemical box model, MECCA (Module Efficiently Calculating the Chemistry of the Atmosphere), was further developed and modified in this investigation. This modified version allowed a better representation of mixing of upwind air masses, and was termed MECCA-MCM-UPWIND. The MECCA-MCM-UPWIND model was used to perform sensitivity studies on the influence of different parameters on O<sub>3</sub> levels in the Jhb-Pta megacity. Possible scenarios to alter or mitigate pollution were also investigated.

## Model description

MECCA is a multi-purpose atmospheric chemistry model that contains chemical mechanisms describing tropospheric and stratospheric chemistry in both the gaseous and aqueous phases.<sup>15</sup> MECCA was originally developed by Sander et al.<sup>15</sup> at the Max-Planck Institute for Chemistry in Mainz, Germany. The full Master of Chemical Mechanism (MCM) was added to the MECCA box model and the MECCA-MCM model was used as described by Butler et al.<sup>16</sup> The MCM v3.1,<sup>17-18</sup> obtained via the website: <http://mcm.leeds.ac.uk/MCM>, is a near-explicit chemical mechanism that describes the detailed gas-phase chemical processes involved in the atmospheric degradation of a series of primarily emitted volatile organic compounds (VOCs). The existing photochemical

box model, MECCA-MCM, was further developed and improved in this study. This model was named MECCA-MCM-UPWIND, which included horizontal and vertical mixing processes, as well as boundary layer height variation. These processes were included to simulate the advection of upwind air masses into the modelling domain, as well as the entrainment from the troposphere through the diurnal mixing layer (ML) height variation, allowing a more detailed simulation of the processes controlling air quality in the Jhb-Pta conurbation.

## Model domain

The model domain was selected to be 100 km x 100 km, which included the Jhb-Pta megacity. The model domain is indicated by the rectangular block in the magnified area in Figure 1.

## Input data

Meteorological data, ML growth, pollutant emissions and pollutant mixing ratios were required as input data for the model. Five stations were identified within the modelling domain and one outside of it, from where data could be obtained. These stations are operated by the Department of Environmental Affairs, the City of Johannesburg and the City of Tshwane. The concentrations of additional species that were not measured at these monitoring stations at the time when data were collected were also required. A three month sampling campaign was therefore conducted from March to May in 2009 to obtain this additional information.<sup>3</sup> The limitations associated with using a three months data set collected during autumn were realised. There are significant differences in meteorological conditions experienced during different seasons. For example, winter has lower temperatures, greater diurnal variations and strong surface inversion layers prevailing in this part of South Africa leading to an increase in concentrations of pollutant atmospheric. However, the main objective for this sampling study was not to conduct a comprehensive measurement campaign, but rather to obtain representative data that could be used for the photochemical modelling. Previously published modelling studies have relied on even shorter measurement campaigns to obtain representative data.<sup>19-21</sup>

Only three of the five measurement stations located within the model domain conducted continuous meteorological measurements during the sampling period, because maintenance was being performed at the other two stations. Meteorological data obtained from these three measurement stations were averaged over a 3-month period (March–May) in 2009 to determine the diurnal variability of meteorological parameters in the Jhb-Pta megacity. According to these data, the average temperature,

atmospheric pressure, and relative humidity for autumn (March–May) in the Jhb-Pta megacity ranged from 12–22 °C, 84.3–84.6 kPa and 40–80%, respectively.

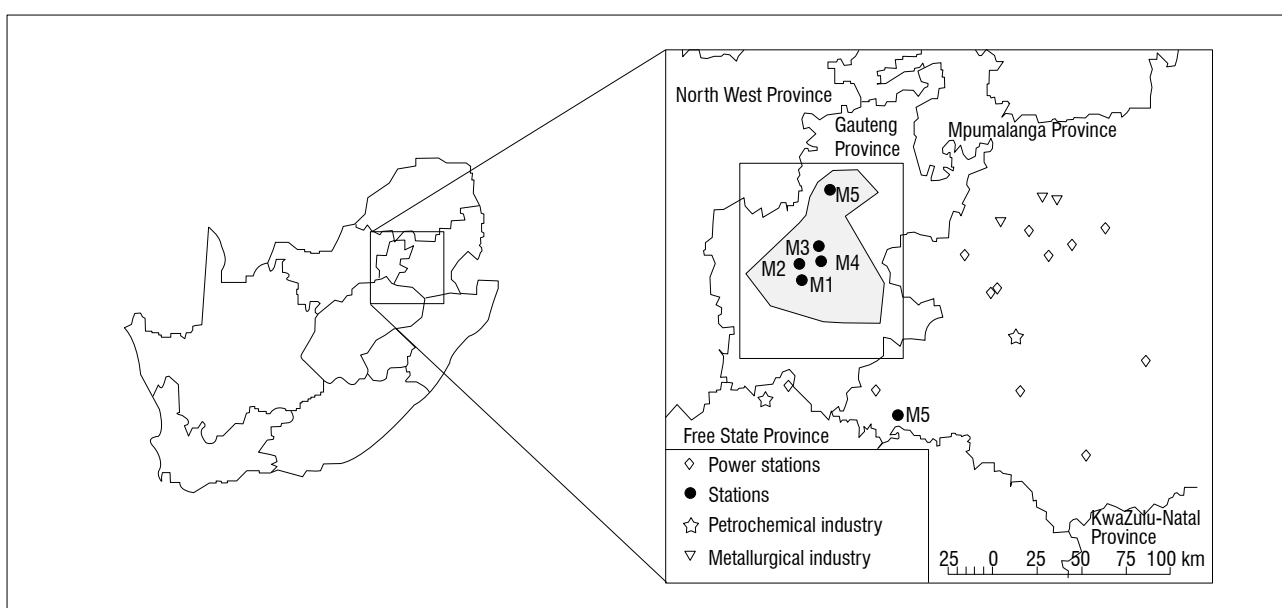
The mixing ratio input data required were the initial, tropospheric and upwind mixing ratios. The initial and tropospheric mixing ratios were obtained from ground-based measurements and published global model runs. The composition of the upwind mixing ratios, i.e. the air mass advected into the Jhb-Pta megacity, were determined by running the MECCA-MCM-UPWIND model with representative data for the Mpumalanga Highveld.

## Mixing layer height

The Air Pollution Model (TAPM) was utilised to simulate surface and free tropospheric meteorological data at a 2.5 km x 2.5 km resolution using synoptic scale meteorology (longitude/latitude grid at 1° grid spacing). The input data for TAPM were obtained from the US National Centre for Atmospheric Research. The average diurnal ML for three months (March–May) for the Jhb-Pta megacity was determined over a five consecutive year period (2004–2009) at a height of 1275 m.

## Emission Inventory

South Africa does not currently have a comprehensive gridded national inventory for sources of atmospheric emissions that are readily available to scientists. This complicated the acquisition of emission data used in the model. Detailed emission data do, however, exist for specific regions. These data sets were mostly obtained through research initiatives (e.g. SAFARI 2000), by metropolitan councils or by private companies on behalf of major industries. An emission inventory was developed for the Mpumalanga Highveld region during the Fund for the Research into Industrial Development Growth and Equity (FRIDGE) campaign in 2004.<sup>22</sup> Although this emission inventory was compiled for the Mpumalanga Highveld region, it also contained emission sources for the Jhb-Pta megacity. These emission sources included residential vehicles, industry, residential activities and biomass combustion as indicated in Table 1. Emissions for the species based on the FRIDGE studies were mainly used in this investigation. The FRIDGE emission inventory was also supplemented with the emission inventories developed for the Vaal Triangle during the Vaal Triangle Airshed Pollution Area (VTAPA) study in 2013<sup>23</sup> and the Gauteng study in 2009<sup>24</sup>, also presented in Table 1.



**Figure 1:** Model domain under investigation, with sampling stations in the Johannesburg-Pretoria (Jhb-Pta) megacity also shown. The Jhb-Pta megacity is indicated by the grey area.

**Table 1:** Emission sources (ton/year) for the Johannesburg-Pretoria megacity based on the Fund for the Research into Industrial Development Growth and Equity (FRIDGE)<sup>21</sup> emission inventory. The Vaal Triangle Airshed Pollution Area (VTAPA)<sup>22</sup> and Gauteng<sup>23</sup> studies are also included as supplementary data. It must also be noted that the use of Pb as an additive in petrol has been phased out since the FRIDGE emission inventory was established

Source types (ton/year)						
FRIDGE campaign						
Species	Vehicle sources	Industrial sources	Biomass burning	Residential/ Domestic fuel burning	Mines and Ash Dumps	Total
CO <sub>2</sub>	5 844 508		405 327	1 546 010		7 795 845
CO	272 579		10 33 122	90 232		362 811
CH <sub>4</sub>	1 211		24 348	3 327		28 886
Other NMHC	33 103			4 117		37 220
Benzene	484	3		139		626
Formaldehyde	292					292
Acetaldehyde	149			164		313
1,3 Butadiene	451					451
NO <sub>x</sub>	66 932	53 934		1 743		122 609
N <sub>2</sub> O						
SO <sub>2</sub>	10 103	20 811		7 877		38 791
Pb	242					242
N <sub>2</sub> O	200			39		239
VTAPA study						
SO <sub>2</sub>	448	247 224	84	826		
NO <sub>x</sub>	16 593	131 778	913	463		
PM <sub>10</sub>	1 068	13 668	1 494	1 836	4675	
Gauteng study						
SO <sub>2</sub>	15 718	23 050	291	13 151		74 850
NO <sub>x</sub>	127 311	108 002	1 505	1 492		264 777
PM <sub>10</sub>	4 663	9 550	2 427	3 796		23 051
Benzene	952	6		73		1 097
Formaldehyde	0.03	584				624
Acetaldehyde	0.06	298		12		332
1,3 Butadiene		906	80	9		1 055

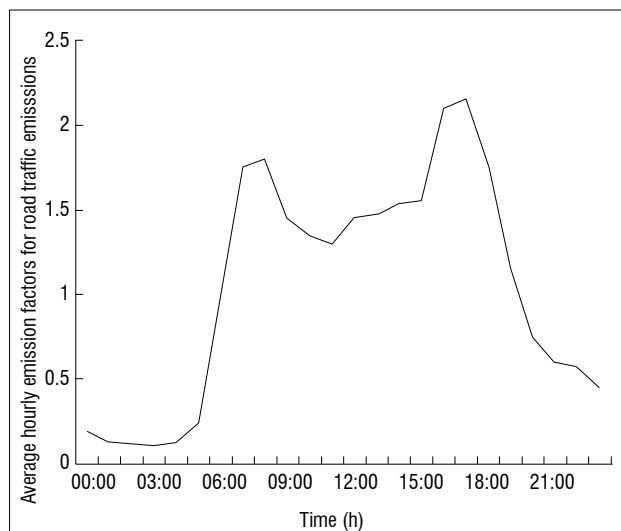
The full MCM chemistry scheme in MECCA-MCM-UPWIND also included numerous non-methane hydrocarbon (NMHC) species. Although some studies have been published reporting BTEX (benzene, toluene, ethyl benzene and xylenes) for very limited sampling areas in South Africa,<sup>3,25</sup> no detailed speciated NMHC data existed for South Africa at the time of the study. Local inventories included benzene, formaldehyde, acetaldehyde and 1,3 butadiene as indicated in Table 1. Emission data (emission rates in molec/cm<sup>2</sup>/s) of NMHC species that were required as input data for the MECCA-MCM-UPWIND model were based on emission mass fractions for European conditions<sup>26</sup> and total carbon emissions (t/year) for Gauteng, which are listed in Table 2. As this was based on European conditions, the estimated NMHC emissions could potentially be an under- or over-estimate.

#### Diurnal variation of traffic emissions

In the Jhb-Pta megacity, traffic is regarded as the major emission source, contributing to approximately 80% of the total nitrogen oxide (NO<sub>x</sub>) emissions.<sup>27</sup> Vehicles also account for approximately 60% of benzene, 100% of formaldehyde, 99% of acetaldehyde and 70% of the total NMHC emissions.<sup>28</sup> Pollutant species from traffic emissions have a significant diurnal variability. Goyns<sup>29</sup> determined the real-world fuel consumption and emissions from passenger vehicles for Johannesburg from 2001 to 2005. According to Goyns<sup>29</sup>, most of the vehicular emissions are emitted during the peak traffic hour periods 06:00–09:00 and 16:00–18:30. The above-mentioned diurnal variability is in general agreement with Menut et al.<sup>30</sup> (Schaberg. 2012. Written communication April 23), who derived averaged hourly emission factors (Figure 2) based on pollutant concentrations



in various cities in Europe by utilising the chemistry transport model CHIMERE. As hourly values were needed for the photochemical box model used in this study, pollutant species were incorporated as hourly emissions in the model, based on the emission factors presented in Figure 2.



**Figure 2:** Diurnal variation of traffic emissions based on factor estimation (adopted from Menut et al.<sup>28</sup>).

Total atmospheric NO and NO<sub>2</sub> concentrations are commonly measured together as NO<sub>x</sub> concentrations. For the model, however, it was required that the emissions of NO and NO<sub>2</sub> species were separated. Various factors influence the NO/NO<sub>2</sub> ratio from raw exhaust gas, e.g. petrol or diesel engines, or exhaust gas from diesel engines after processing by a catalytic converter.<sup>31</sup> The NO<sub>2</sub> fraction in NO<sub>x</sub> varies between 5% and 25% by mass, depending on the type of vehicle and fuel.<sup>31</sup> The European Monitoring and Evaluation Programme (EMEP) of the European Environmental Agency (EAA) compiled an air pollution emission inventory guidebook (EMEP/EEA), which provides guidance on estimating emissions from both anthropogenic or natural emission sources.<sup>32</sup> Based on the EMEP/EEA guidebook, it was assumed in the model that the NO/NO<sub>2</sub> ratio for raw exhaust gas vehicle emissions was 80% NO and 20% mass fraction NO<sub>2</sub>.

The Euro fuel specification is defined as the acceptable limits for exhaust emissions of new vehicles in Europe in order to meet their stringent emission regulations. The 'criteria' atmospheric pollutants are regulated through these standards. South Africa has also adopted this standard classification reference method. The different Euro specifications are an indication of when they came into force, i.e. Euro-0, pre-1993; Euro-1 in 1993; Euro-2 in 1996; Euro-3 in 2000, Euro-4 in 2005 and Euro-6 in 2014. According to Goyns<sup>29</sup>, the private passenger vehicle fleet in the Jhb-Pta megacity from 2001 to 2005 consisted of approximately 45% Euro-0 type petrol vehicles, 10% Euro-2 diesel vehicles, 30% Euro-2 petrol vehicles and 15% Euro-3 petrol vehicles. The percentages of each pollutant species emitted from the different Euro classified vehicles that was used in this study are listed in Table 3.

**Table 2:** NMHC (non-methane hydrocarbon) emissions (ton/year) based on the mass fraction (adopted from Derwent et al.<sup>24</sup>)

Species	Name	ton/year	Species	Name	ton/year
Isobutene1-butene	C <sub>4</sub> H <sub>10</sub>	151	Butane	C <sub>4</sub> H <sub>10</sub>	1312
mp-Xylene	C <sub>8</sub> H <sub>10</sub>	939	trans-2-Butene	C <sub>4</sub> H <sub>8</sub>	288
o-Xylene	C <sub>8</sub> H <sub>10</sub>	813	cis-2-Butene	C <sub>4</sub> H <sub>8</sub>	288
1,2,3-Trimethyl-benzene	C <sub>9</sub> H <sub>12</sub>	166	1-Pentene	C <sub>5</sub> H <sub>10</sub>	99
Acetone	C <sub>3</sub> H <sub>6</sub> O	785	n-Pentane	C <sub>5</sub> H <sub>12</sub>	634
Ethanol	C <sub>2</sub> H <sub>6</sub> O	4474	Isoprene	C <sub>5</sub> H <sub>8</sub>	1
Ethylene	C <sub>2</sub> H <sub>4</sub>	1168	2-Methyl-2-butene	C <sub>5</sub> H <sub>10</sub>	83
Toluene	C <sub>7</sub> H <sub>8</sub>	2149	n-Hexane	C <sub>6</sub> H <sub>14</sub>	420
Propylene	C <sub>3</sub> H <sub>6</sub>	50	2-Methyl-heptane	C <sub>6</sub> H <sub>18</sub>	498
Propane	C <sub>3</sub> H <sub>8</sub>	164	Cyclohexane	C <sub>6</sub> H <sub>12</sub>	1
Isobutene	C <sub>4</sub> H <sub>10</sub>	1312	2-Methyl-hexane	C <sub>7</sub> H <sub>16</sub>	167
n-Octane	C <sub>8</sub> H <sub>18</sub>	89	n-Heptane	C <sub>7</sub> H <sub>16</sub>	108
Ethyl-benzene	C <sub>8</sub> H <sub>10</sub>	409			

**Table 3:** Input data for the Johannesburg-Pretoria (Jhb-Pta) megacity model run

Euro-0 petrol (42% of total Jhb-Pta megacity passenger fleet)	Euro-2 diesel (10% of total Jhb-Pta megacity passenger fleet)	Euro-2 petrol (30% of total Jhb-Pta megacity passenger fleet)	Euro-3 petrol (15% of total Jhb-Pta megacity passenger fleet)
40% CO <sub>2</sub>	19% CO <sub>2</sub>	27% CO <sub>2</sub>	15% CO <sub>2</sub>
88% CO		10% CO	2% CO
93% HC		5% HC	0.5% HC
82% NO <sub>x</sub>	14% NO <sub>x</sub>	3% NO <sub>x</sub>	0.4% NO <sub>x</sub>

**Table 4:** Vehicle emissions for Johannesburg-Pretoria megacity according to the Euro classification<sup>29</sup>

Species	Formula	Initial mixing ratio (ppb)	Tropospheric mixing ratio (ppb)	Upwind mixing ratio (ppb) from Upwind model run	Emissions (molec/cm <sup>2</sup> /s)
Nitrogen monoxide	NO	20	0.1	40	1.2x10 <sup>11†</sup>
Nitrogen dioxide	NO <sub>2</sub>	14	1	38	1.0x10 <sup>10†</sup>
Peroxyacetyl nitrate	PAN	2	0.02	0.85	-
Ozone	O <sub>3</sub>	7	65	19	-
Carbon monoxide	CO	2700	75	2982	9.1x10 <sup>12†</sup>
Carbon dioxide	CO <sub>2</sub>		345 <sup>‡</sup>		6.7x10 <sup>11</sup>
Methane	CH <sub>4</sub>	1800	175 <sup>‡</sup>	6260	1.2x10 <sup>11</sup>
Ethane	C <sub>2</sub> H <sub>6</sub>		0.4	5.8	12.1x10 <sup>8</sup>
Ethylene	C <sub>2</sub> H <sub>4</sub>				1.7x10 <sup>9†</sup>
Formaldehyde	CH <sub>2</sub> O				6.6x10 <sup>8†</sup>
Propylene	C <sub>3</sub> H <sub>6</sub>				1.4x10 <sup>9†</sup>
Propane	C <sub>3</sub> H <sub>8</sub>			0.64	1.3x10 <sup>9†</sup>
Isobutene 1-butene	C <sub>4</sub> H <sub>10</sub>				5.4x10 <sup>8†</sup>
Butane	C <sub>4</sub> H <sub>10</sub>				1.9x10 <sup>9†</sup>
Trans-2-Butene	C <sub>4</sub> H <sub>8</sub>				4.3x10 <sup>8†</sup>
Cis-2-Butene	C <sub>4</sub> H <sub>8</sub>				4.3x10 <sup>8†</sup>
1-pentene	C <sub>5</sub> H <sub>10</sub>				1.2x10 <sup>8†</sup>
N-Pentane	C <sub>5</sub> H <sub>12</sub>				7.7x10 <sup>8†</sup>
2-Methyl-2-butene	C <sub>5</sub> H <sub>10</sub>				1.0x10 <sup>8†</sup>
N-Hexane	C <sub>6</sub> H <sub>14</sub>				5.8x10 <sup>8†</sup>
Benzene	C <sub>6</sub> H <sub>6</sub>	3	0.01 <sup>†</sup>	0.62	5.5x10 <sup>8†</sup>
Cyclohexane	C <sub>6</sub> H <sub>12</sub>				4.9x10 <sup>5†</sup>
2-Methyl-hexane	C <sub>7</sub> H <sub>16</sub>				1.4x10 <sup>8†</sup>
N-Heptane	C <sub>7</sub> H <sub>16</sub>				9.3x10 <sup>7†</sup>
Toluene	C <sub>7</sub> H <sub>8</sub>	3	0.01	0.8	9.8x10 <sup>8†</sup>
2-Methyl-heptane	C <sub>8</sub> H <sub>18</sub>				3.7x10 <sup>8†</sup>
N-Octane	C <sub>8</sub> H <sub>18</sub>				6.7x10 <sup>7†</sup>
Ethyl-benzene	C <sub>8</sub> H <sub>10</sub>	4	0.01	0.23	2.7x10 <sup>8†</sup>
Mp-Xylenes	C <sub>8</sub> H <sub>10</sub>	4	0.01	0.42	9.9x10 <sup>7†</sup>
O-Xylene	C <sub>8</sub> H <sub>10</sub>	4	0.01	0.41	2.4x10 <sup>8†</sup>
1,2,3-Trimethyl-benzene	C <sub>9</sub> H <sub>12</sub>				9.8x10 <sup>8†</sup>
Acetone	C <sub>3</sub> H <sub>6</sub> O				7.0x10 <sup>12†</sup>
Acetylene	C <sub>2</sub> H <sub>2</sub>		0.04		1.9x10 <sup>10†</sup>
Ethanol	C <sub>2</sub> H <sub>6</sub> O				4.0x10 <sup>9†</sup>
1,3-butaine	C <sub>4</sub> H <sub>6</sub>				1.0x10 <sup>8†</sup>

<sup>†</sup> Emissions emitted hourly in the model. Listed values present the emission at the start of the model run.

<sup>‡</sup> Mixing ratio in part per million (ppm)

As indicated in Table 3, Euro-0 petrol vehicles have the highest emissions and produce 40% CO<sub>2</sub>, 88% CO, 93% hydrocarbon (HC) and 82% NO<sub>x</sub> emissions, whereas Euro-3 petrol vehicles contribute much less, with 15% CO<sub>2</sub>, 2% CO, 0.5% HC and 0.4% NO<sub>x</sub> emissions. These derived emissions excluded heavy-duty vehicles and minibus taxis, which also contribute significantly to vehicular emissions in the Jhb-Pta megacity. In this study, the worst case scenario was assumed in the model base case run, i.e. that all vehicles in the Jhb-Pta megacity were Euro-0 type. During the sensitivity analysis, the impact of the conversion of all vehicles to Euro-3 type on pollutant species in the atmosphere was also investigated.

### Initial and tropospheric mixing ratios

The averaged initial mixing ratios of most of the criteria pollutants (NO, NO<sub>2</sub>, O<sub>3</sub> and BTEX) were obtained from data collected at the stations indicated in Figure 1, from literature,<sup>15,33,34</sup> as well as from data collected during a three-month (March–May 2010) monitoring campaign.<sup>3</sup> These average initial mixing ratios are listed in Table 4. Measurements conducted at midnight (00:00) were used as the initial mixing ratio values in the MECCA-MCM-UPWIND model run, because the start time of the model is 00:00.

No data were available for most of the species in the Jhb-Pta megacity as a function of height above ground level within the troposphere and therefore these mixing ratios were estimated, either through model simulations or obtained from literature. Estimated tropospheric mixing ratios for species were obtained from the global chemistry-climate model (EMAC).<sup>15</sup> Output files from EMAC were generated for each day for all the major pollutant species (i.e. CO, CH<sub>4</sub>, PAN, NO, NO<sub>2</sub>, SO<sub>2</sub>) and used as input into the MECCA-MCM-UPWIND model. With the exception of CO and CH<sub>4</sub>, the other most important tropospheric organic trace gases are ethane (C<sub>2</sub>H<sub>6</sub>) and ethyne (C<sub>2</sub>H<sub>2</sub>).<sup>15,33</sup> C<sub>2</sub>H<sub>6</sub> is the second most abundant HC in the atmosphere and has a lifetime of approximately two months,<sup>35</sup> whereas C<sub>2</sub>H<sub>2</sub> has an estimated lifetime of 2–4 weeks in the atmosphere.<sup>36</sup> No data were available for these species in the Jhb-Pta megacity, so the global mixing ratios in the southern hemisphere of 0.04 ppb and 0.044 ppb for C<sub>2</sub>H<sub>6</sub> and C<sub>2</sub>H<sub>2</sub>, respectively, were used.<sup>32</sup> The global mixing ratios for CO<sub>2</sub>, CO and CH<sub>4</sub> were also used.<sup>37</sup> Most vertical tropospheric O<sub>3</sub> profiling studies in southern Africa have been conducted during sampling campaigns, e.g. SAFARI 2000. Vertical tropospheric O<sub>3</sub> measured at Irene (Pretoria) was reported to be 60–65 ppb, between 2 km and 2.5 km above ground level.<sup>38,39</sup> Concentrations of species such as O<sub>2</sub>, N<sub>2</sub> and CO<sub>2</sub> were fixed throughout the model runs.

### Upwind mixing ratios

This study also includes a representation of advection of air masses from the highly polluted Mpumalanga Highveld into the Jhb-Pta megacity because of its proximity to the Jhb-Pta megacity. The composition of the air masses entering the Jhb-Pta megacity was simulated using MECCA-MCM-UPWIND making use of input data representative of the Mpumalanga Highveld. A total of 5 air quality monitoring stations within the Mpumalanga Highveld were used to obtain initial mixing ratios for this model run.<sup>3</sup> The average wind speed used was 1.7 m/s as determined at four of these air quality monitoring stations. Model spin-up time of 24 h was performed in order to provide realistic initial concentrations for the second 24 h. The model results averaged over the second 24 h period were used as the upwind mixing ratio inputs for the Jhb-Pta megacity base-case and sensitivity runs. The same ML height utilised for the Jhb-Pta megacity was used for this model run, as the Mpumalanga Highveld is situated at approximately the same height above sea level. Recently Korhonen et al.<sup>40</sup> also indicated that the ML over the Mpumalanga Highveld and Jhb-Pta are similar.

Industrial activities are the main emission sources in the Mpumalanga Highveld. These industries mostly operate 24 h per day, so there is no diurnal variability in the emissions.<sup>41</sup> Therefore it was not necessary to specify a temporal profile of emissions in the model run. Emission data for the Mpumalanga Highveld obtained from the FRIDGE study were utilised.

According to literature, the initial NO/NO<sub>2</sub> ratio on release from a coal-fired power station into the atmosphere is about 95/5 by mass, while

the release from vehicles is 80/20 by mass. However, traffic emissions in the Mpumalanga Highveld are much lower compared to industrial emissions. Therefore, the assumption to use the ratio of NO/NO<sub>2</sub> from the Highveld as 95% NO and 5% mass fraction NO<sub>2</sub> was made.<sup>42</sup>

The mixing ratios of all the pollutant species measured at the air quality monitoring stations in the Mpumalanga Highveld were used as initial mixing ratios for the Highveld model run. Average diurnal concentrations for the three months (March–May) were used. For this model run, the upwind mixing of species in the Mpumalanga Highveld was not included as the simulation focussed on the aging of air mass transported between the Mpumalanga Highveld and the Jhb-Pta megacity. Therefore, the upwind mixing ratio of the species within the Mpumalanga Highveld was the same as the initial mixing ratios for all the simulations. The result from this Mpumalanga Highveld model run, which was used as input for the Jhb-Pta megacity, are summarised in Table 4.

## Results and Discussion

### Base case run

A base case model run using the input data, as previously discussed, was used to compare predicted pollutant mixing ratios to measured mixing ratios obtained during the sample campaign. The averaged wind speed of 1.7 m/s based on the metrological data was used in the Jhb-Pta model runs. Figure 3 presents the comparison between the measured and modelled results for NO, NO<sub>2</sub> and O<sub>3</sub> mixing ratios.

The modelled diurnal NO pattern was in general in good agreement with the measured NO. The diurnal variation in the modelled NO can be attributed to changes in emissions, transport, chemistry and the ML height in the box defined for the model. The peak of NO in the morning and evening was well reproduced by MECCA-MCM-UPWIND. This increased NO mixing ratio can be ascribed to increased emissions within the model. The measured data also showed the same tendency with NO usually peaking in the morning because of increases in emissions from vehicles and household combustion, in conjunction with a relatively low boundary layer height. The early morning peak between 6:00 and 9:00 coincided with the time that commuters travel to work.<sup>3</sup> The second peak observed between 16:00 and 21:00 for both the measured and modelled data can be attributed to traffic emissions and household combustion (space heating and cooking).<sup>43</sup> A decrease in NO emissions was observed for the measured data after 19:00. This is the result of a significant decrease in NO sources. However, the model results indicated an increase in NO mixing ratios. This can be explained by emission and ML effects in the model. When the ML decreases within the model, it stops the vertical mixing from above the ML, which results in the accumulation of NO because of the decrease in volume and emissions emitted into the model.

It seems that the modelled NO<sub>2</sub> mixing ratios were somewhat over-predicted by MECCA-MCM-UPWIND, although it was in relatively good agreement with the diurnal variability determined with the measured NO<sub>2</sub> concentrations. Because 20% of NO<sub>x</sub> emitted from vehicles are NO<sub>2</sub>, diurnal changes of traffic emissions will also influence the trend observed. The first NO<sub>2</sub> peak can be explained by the increased traffic emissions resulting in higher NO<sub>2</sub> mixing ratios and decrease as the ML height increases, which results in NO and NO<sub>2</sub> mixing ratios decreasing, which implies that less NO is available to chemically transform to NO<sub>2</sub>. This leads to a decrease in NO<sub>2</sub> mixing ratios. The second observed peak later in the day can be related to the slowed photolysis rate of NO<sub>2</sub>, resulting in the accumulation of NO<sub>2</sub>.

The diurnal cycle of O<sub>3</sub> was well predicted by MECCA-MCM-UPWIND and is in good agreement with measured concentrations, although some differences between the measured and modelled data were observed. MECCA-MCM-UPWIND simulated very low mixing ratios of O<sub>3</sub> during the early morning hours and evening hours with a slight over-prediction during the daytime mixing ratio of O<sub>3</sub>. O<sub>3</sub> is accumulated through photochemical oxidation processes throughout the day until dusk. At this time, the photochemical processes slow down because of to the lower levels of solar radiation. The ML height decrease also results in no entrainment from above. The almost zero O<sub>3</sub> levels during night-time for the model run can be attributed to the NO<sub>x</sub>-titration effect within the model.

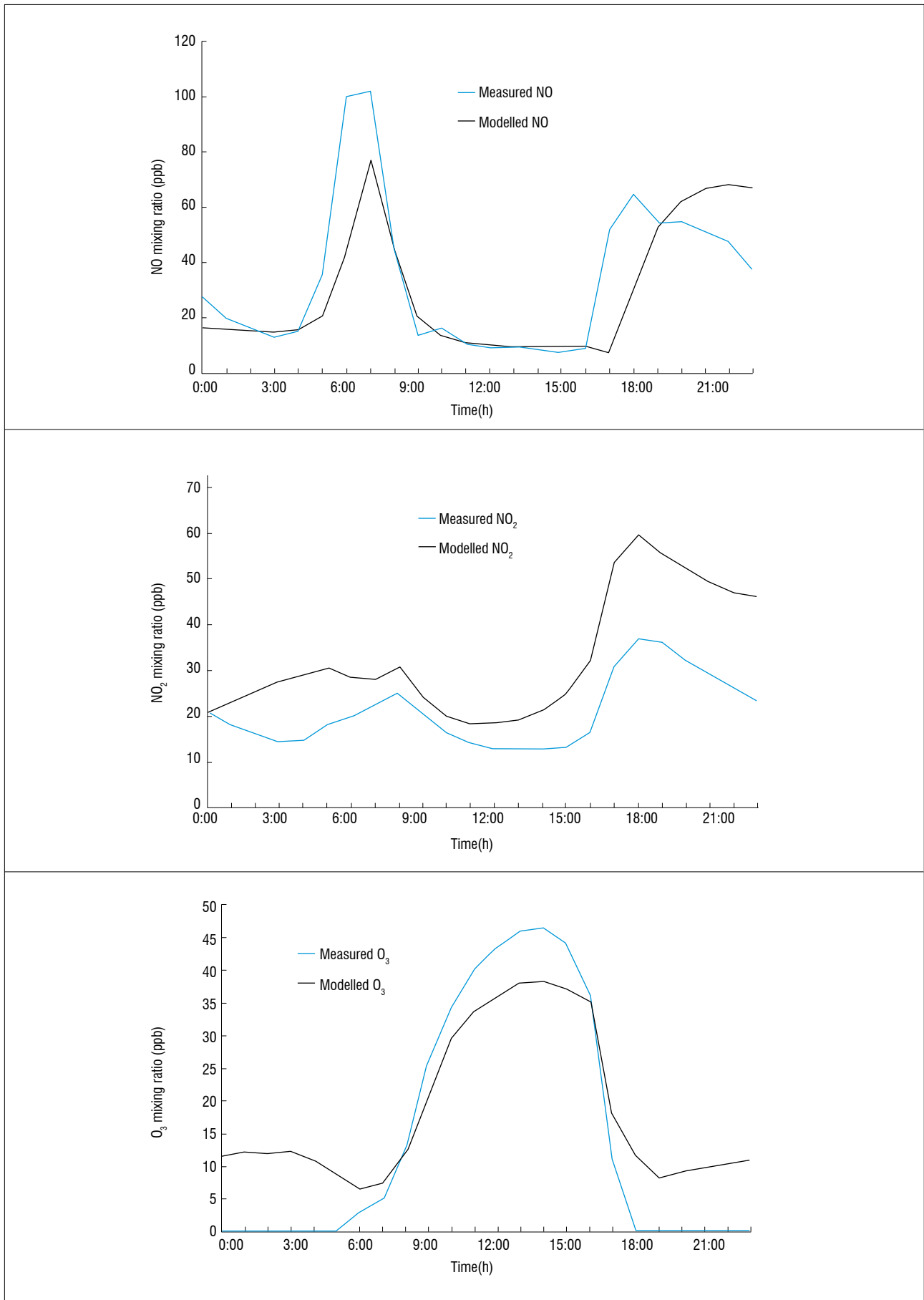


Figure 3: Modelled and measured diurnal mixing ratios for NO, NO<sub>2</sub> and O<sub>3</sub>.



Model results presented later in the paper (Figure 5) indicate that there is enough NO in the model to completely convert all the O<sub>3</sub> to NO<sub>2</sub>. NOx titration can drive O<sub>3</sub> concentrations in urban areas to less than 1 ppb.<sup>44</sup> The discrepancy between the measured and modelled results can be considered as a possible limitation of the model. However, other modelling studies, e.g. Stockwell et al.<sup>45</sup>, have emphasised the difficulty of correctly representing night-time chemistry in models.

#### Sensitivity of O<sub>3</sub> production to changes in wind speed

In this model run, the sensitivity of the Jhb-Pta megacity to the rate at which external pollution is advected into the megacity was examined. Although O<sub>3</sub> concentrations in the atmosphere are influenced by changes in meteorological conditions, such as ML height, temperature and wind speed, in this scenario only wind speed was varied. The influence of no wind (0 m/s), medium wind speed (1.7 m/s) and high wind speed

(3.4 m/s) from the Mpumalanga Highveld was calculated with the MECCA-MCM-UPWIND model. The results presented in Figure 4 indicate that lower wind speeds lead to increases in O<sub>3</sub> mixing ratios in the Jhb-Pta megacity. The results show that wind speed plays an important role in the processes that control O<sub>3</sub> concentrations in the atmosphere. According to the model, O<sub>3</sub> levels are 12 ppb higher when there is no wind present compared to medium wind speed, while O<sub>3</sub> mixing ratios are 22 ppb lower during medium wind speed conditions compared to higher wind speeds.

When the upwind air mass enters the Jhb-Pta megacity, it is instantaneously mixed in the model with the air in the Jhb-Pta megacity. This implies that if the wind speed doubles and other conditions remain equal, more NOx is transported from the Mpumalanga Highveld into the model, which can lead to the titration of O<sub>3</sub> within the Jhb-Pta megacity as is evident from Figure 5.

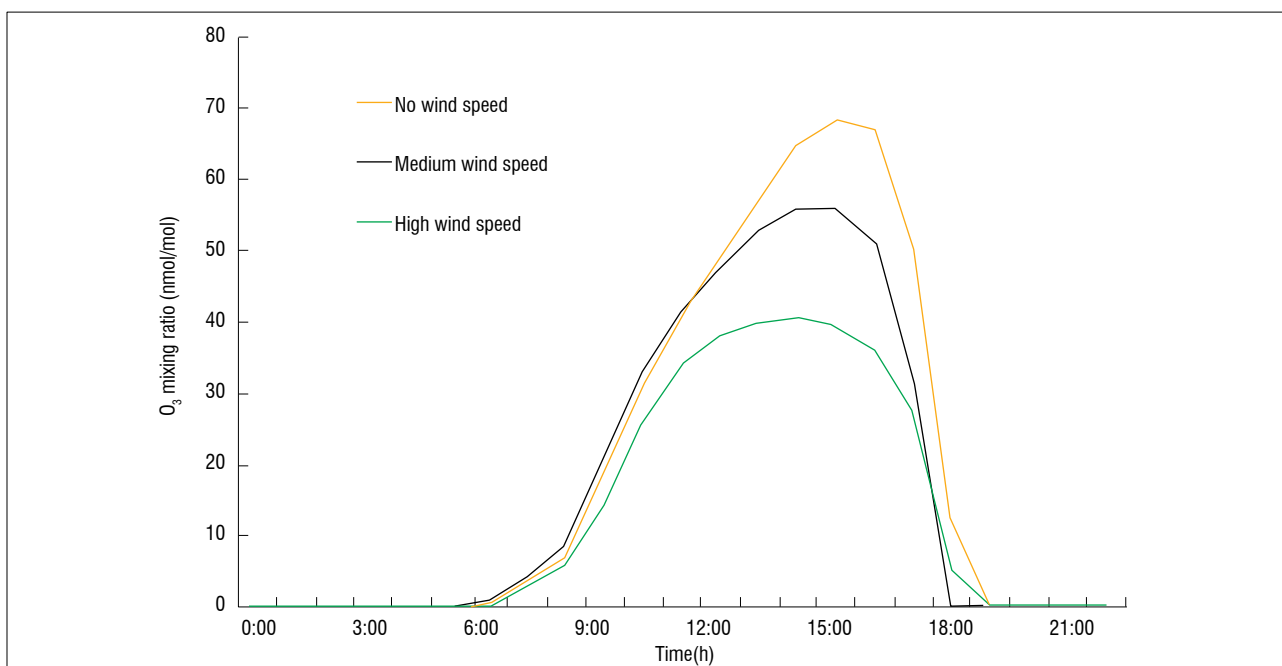


Figure 4: Sensitivity of the model to O<sub>3</sub> mixing ratios with changes in wind speed.

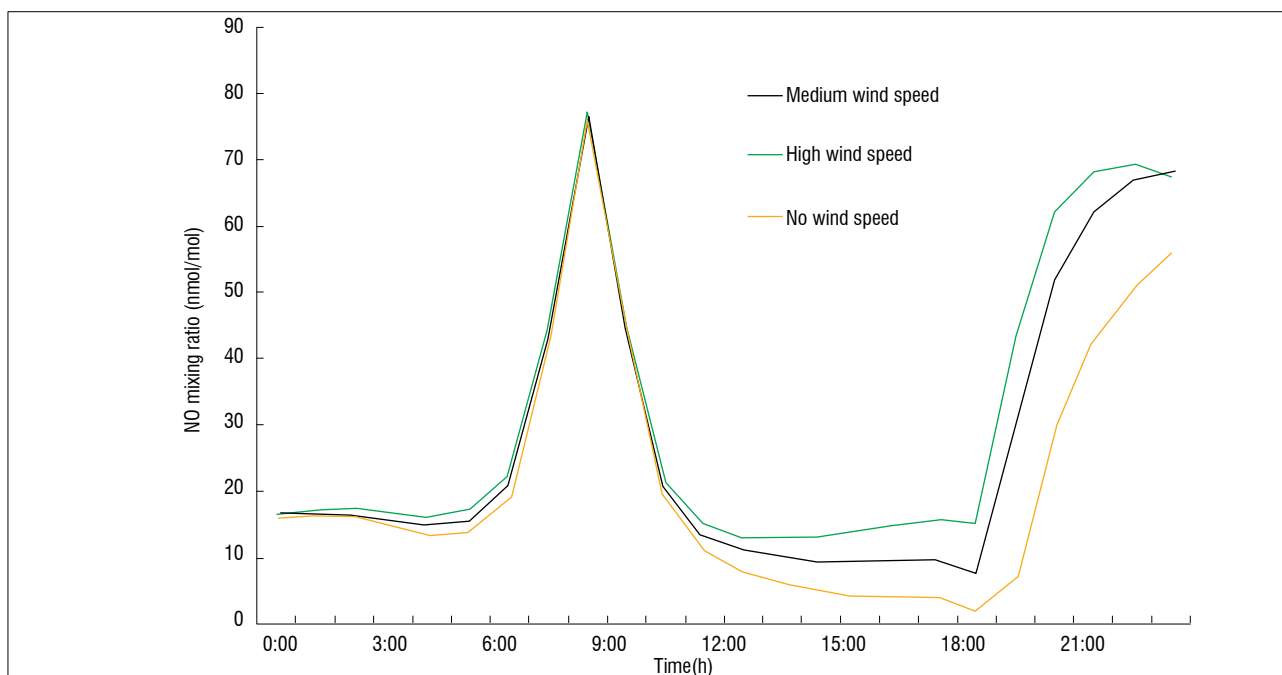
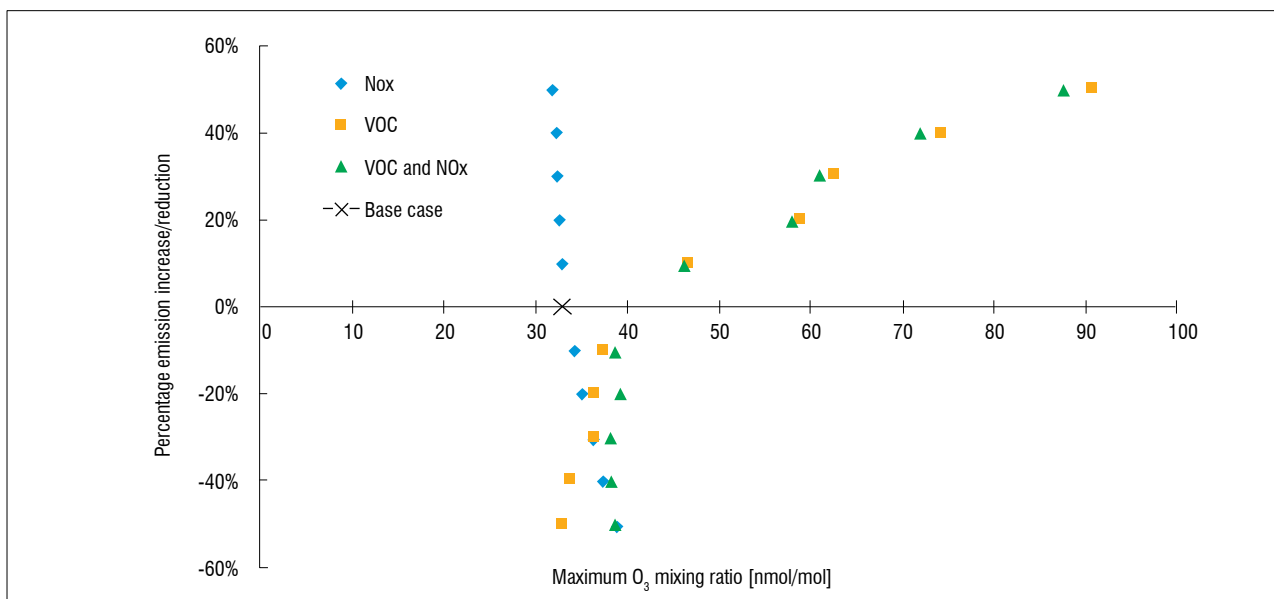
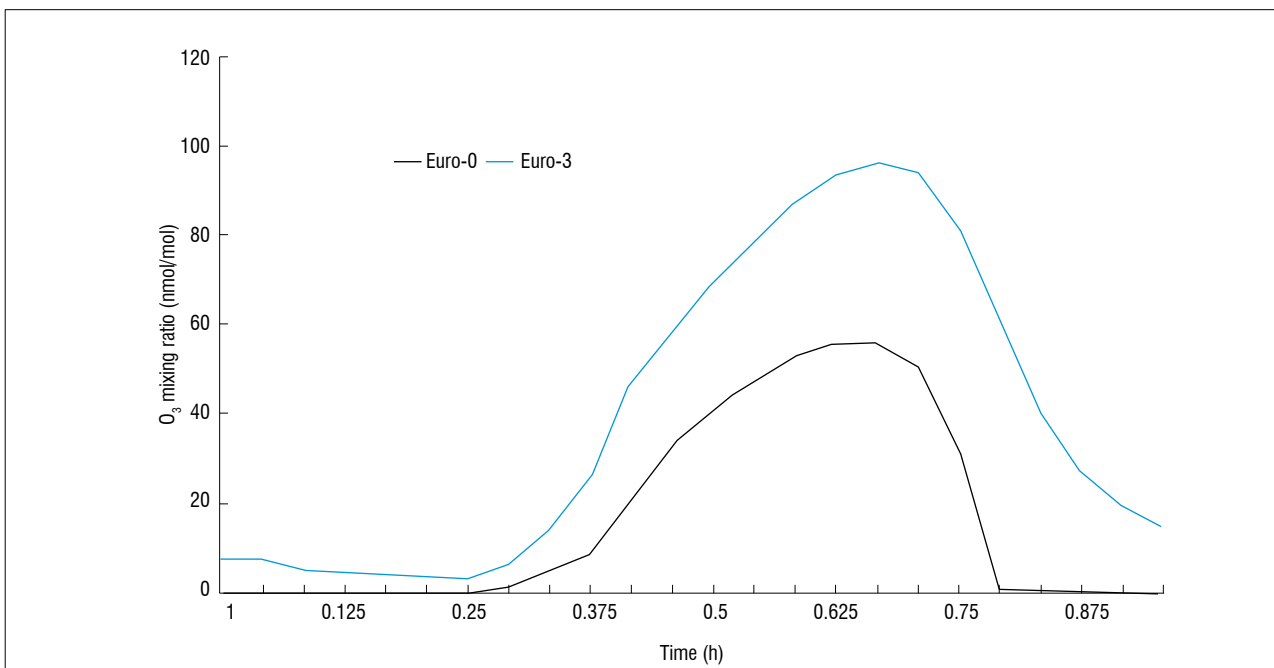


Figure 5: Sensitivity of the model to NO mixing ratios with changes in wind speed.



**Figure 6:** The change in the maximum  $O_3$  mixing ratios associated with each percentage increase/decrease of NOx and VOCs (volatile organic compounds) emissions.



**Figure 7:** Comparison of  $O_3$  production in the Johannesburg-Pretoria megacity by changing from Euro-0 to Euro-3 classified vehicles.

Lower wind speeds implicate longer reaction times, assuming all other conditions stay the same. Additionally,  $O_3$  is a secondary pollutant whose concentration will increase with time as the precursor species react. It is evident from Figure 5 that NO levels increase at higher wind speeds from the Mpumalanga Highveld. The same effect was observed for  $O_3$ . When a lower concentration of  $O_3$  is advected into the Jhb-Pta megacity at twice the rate, the concentrations of  $O_3$  within the Jhb-Pta megacity will decrease. Increasing the  $O_3$ -titration effect results in decreased  $O_3$  levels in the Jhb-Pta megacity. In stagnant conditions where no upwind air is advected into the Jhb-Pta megacity, the concentration of the pollutants in the Jhb-Pta megacity is only influenced by the initial concentrations, emission levels, vertical mixing and chemistry within the megacity. Under these conditions, the local emissions and vertical mixing tend to contribute significantly to the photochemistry for the formation of  $O_3$ .

### Sensitivity of $O_3$ production to changes in Jhb-Pta megacity emissions

In order to calculate the effect of emissions of VOCs and NOx in the Jhb-Pta megacity on the photochemical production of  $O_3$  occurring within the megacity, six case studies were conducted. These case studies were conducted for the reduction and increase in VOC and NOx concentrations separately, as well as in conjunction with each other. In order to avoid potential nonlinearities associated with changing the chemical regime (e.g. NOx-limited or VOC-limited), increments of 10–50% were used. The maximum  $O_3$  mixing ratio resulting from the percentage change for each case study was plotted (Figure 6). The  $O_3$  base case concentration is indicated by the black cross at 56 ppb.

The relationship between NO<sub>x</sub> and VOC concentrations for the production of O<sub>3</sub> can be illustrated by means of isopleth plots.<sup>46</sup> On these isopleth plots, certain chemical regimes can be identified, i.e. areas where O<sub>3</sub> formation depends either on NO<sub>x</sub> or VOC concentrations. The results obtained from the incremental changes in Figure 6 can be related to these O<sub>3</sub> isopleth plots. An increase in O<sub>3</sub> production is observed when VOC emissions are increased, while O<sub>3</sub> levels decrease when NO<sub>x</sub> emissions are increased. Reducing NO<sub>x</sub> emissions while keeping VOC emissions constant will therefore result in an increase in O<sub>3</sub> concentrations. The changes of O<sub>3</sub> mixing ratios resulting from the increase in NO<sub>x</sub> emissions are lower than the increase of O<sub>3</sub> levels observed when NO<sub>x</sub> emissions are reduced. The changes observed with the model can be related to the isopleths in the VOC-sensitive regime. Chemically, this observation can be explained by the reaction of the HO• radical with NO<sub>2</sub> producing HNO<sub>3</sub> that is dominant at high concentrations of NO<sub>x</sub> and suppresses the concentration of the HO• radical. The production of HO<sub>2</sub>• and RO<sub>2</sub>• radicals, which are produced by the oxidation of VOCs with HO• radicals, is slowed down and results in less O<sub>3</sub> production. Regarding the O<sub>3</sub> isopleths, the results shown in Figure 6 indicate that the Jhb-Pta megacity is a VOC-limited (or NO<sub>x</sub>-saturated) regime. Therefore, according to the model, O<sub>3</sub> production in the Jhb-Pta megacity will be reduced most significantly when the VOC emissions in Jhb-Pta megacity are reduced.

#### Effect of vehicle fleet emissions on O<sub>3</sub> production

The fuel strategy of the South African government is to only have vehicles on the roads classified as Euro-4 and higher emissions standards in the future.<sup>47</sup> Therefore, in this case study, the effect of reducing vehicular emissions in the Jhb-Pta megacity on the production of O<sub>3</sub> was also investigated. Based on the values of emissions associated with each Euro-classified vehicle listed in Table 3, a model sensitivity run was performed in which emissions of VOCs, NO<sub>x</sub> and CO were reduced from Euro-0 to Euro-3 vehicles. The results are presented in Figure 7. As is evident, a significant increase in O<sub>3</sub> production of approximately 23 ppb is observed. This is consistent with other modelled sensitivity studies of traffic emissions that also predict that future urban O<sub>3</sub> concentrations will increase in many cities by 2050 because of the reduction in NO<sub>x</sub> titration of O<sub>3</sub>, despite the implementation of O<sub>3</sub> control regulations.<sup>48</sup> The increase of O<sub>3</sub> is also consistent with the NO<sub>x</sub> saturated regime as indicated in Figure 6, which indicated that VOCs should be reduced concurrently with NO<sub>x</sub> in order to avoid the increase in O<sub>3</sub> levels.

## Conclusions

Compared to measured data, MECCA-MCM-UPWIND was able to predict the diurnal variability of NO, NO<sub>2</sub> and O<sub>3</sub> within the Jhb-Pta megacity relatively well. Daytime chemistry was especially well simulated, while small under-predictions were calculated for night-time chemistry. Some of the uncertainties in the model can possibly be attributed to the emission inventories utilised and the deficiency of well-defined NMHC data for the Jhb-Pta megacity.

Sensitivity analyses showed that O<sub>3</sub> mixing ratios decreased with increasing wind speed and increased with decreasing wind speed within the Jhb-Pta megacity. This indicated that the Mpumalanga Highveld could potentially be a major source of NO<sub>x</sub> in the Jhb-Pta megacity. This implies that if the air quality for the larger surrounding area improves, the concentration of the secondary pollutant O<sub>3</sub> will increase in the Jhb-Pta megacity.

The influence of NO<sub>x</sub> and VOC concentrations on O<sub>3</sub> formation in the Jhb-Pta megacity indicated that the Jhb-Pta megacity is a VOC-limited (or NO<sub>x</sub>-saturated) regime. Therefore, O<sub>3</sub> levels in the Jhb-Pta megacity will be more effectively reduced if VOC emissions decrease. A reduction of NO<sub>x</sub> emissions will lead to an increase in O<sub>3</sub> because of a decrease in titration through the reaction with NO. The same effect was observed in various cities worldwide where O<sub>3</sub> levels increased when NO<sub>x</sub> emissions were reduced during emission control strategies. The increase of O<sub>3</sub> can be avoided if VOCs are also reduced.

The effect of reducing vehicular emissions in the Jhb-Pta megacity on the instantaneous production of O<sub>3</sub> was also investigated. A significant increase of approximately 23 ppb O<sub>3</sub> production was observed when

emissions of VOCs, NO<sub>x</sub> and CO were reduced from Euro-0 to Euro-3 vehicles. This is consistent with other modelled sensitivity studies of traffic emissions, which also predict that future urban O<sub>3</sub> concentrations will increase in many cities by 2050 as a result of the reduction in NO<sub>x</sub> titration of O<sub>3</sub>, despite the implementation of O<sub>3</sub> control regulations.

The limitation associated with comparison of the modelled data with a three month measured data set was realised. It is recommended in future that modelled data be compared to a complete 1-year ground measurement data set, which will account for different meteorological conditions occurring in each of the seasons. Future work should also aim at improving the model to reduce over- and under-prediction of concentrations of species. Expanded and accurate emission inventories for South Africa will be pivotal in improving modelled data sets. It is also important that changes in the vehicular fleet should be reflected in improved emission inventories.

## Acknowledgments

The authors acknowledge Sasol Research & Development and the National Research Foundation for their financial support.

## Authors' contributions

A.S.M.L., T.M.B., J.P.B. and P.G.v.Z. were the main investigators in this study. A.S.M.L., T.M.B., J.P.B. and P.G.v.Z. were project leaders of the study and wrote the manuscript. A.S.M.L. conducted this study as part of her PhD and performed most of the experimental work. P.G.v.Z. and J.P.B. were also study leaders of this PhD study, while T.M.B. was an assistant study leader. G.D.F. and M.G.L. made conceptual and logistical contributions.

## References

- Molina MJ, Molina LT. Megacities and atmospheric pollution. *J Air Wast Manage.* 2004;54:644–680. <http://dx.doi.org/10.1080/10473289.2004.10470936>
- Lourens ASM, Beukes JP, Van Zyl PG, Fourie GD, Burger JW, Pienaar JJ, et al. Spatial and temporal assessment of gaseous pollutants in the Mpumalanga Highveld of South Africa. *S Afr J Sci.* 2011;107(1/2) Art. #269, 8 pages. <http://dx.doi.org/10.4102/sajs.v107i1/2.269>
- Lourens ASM, Butler TM, Beukes JP, Van Zyl PG, Beirle S, Wagner T, et al. Re-evaluating the NO<sub>2</sub> hotspot over the South African Highveld. *S Afr J Sci.* 2012;108(11/12) Art. #1146, 6 pages. <http://dx.doi.org/10.4102/sajs.v108i11/12.1146>
- National Environmental Management: Air Quality Act No. 39 of 2004 section 63. *Government Gazette.* 2005 Feb 24; Vol. 476, No. 27318. Available from: [https://www.environment.gov.za/sites/default/files/legislations/nema\\_amendment\\_act39.pdf](https://www.environment.gov.za/sites/default/files/legislations/nema_amendment_act39.pdf)
- Jin S, Demerjian KL. A photochemical box model for urban air quality study. *Atmos Environ.* 1993;278(1):371–387. [http://dx.doi.org/10.1016/0957-1272\(93\)90015-X](http://dx.doi.org/10.1016/0957-1272(93)90015-X)
- Schere KL, Demerjian KL. User guide for the photochemical box model (PBM). Research Triangle Park, NC: Environmental Sciences Research Laboratory; 1984.
- Butler TM, Lawrence MG, Taraborrelli D, Lelieveld J. Tagged ozone production potential (TOPP) of volatile organic compounds. *Atmos Environ.* 2011;45:4082–4090. <http://dx.doi.org/10.1016/j.atmosenv.2011.03.040>
- Kleinman LI. Seasonal dependence of boundary layer peroxide concentration: The low and high NO<sub>x</sub> regimes. *J Geophys Res.* 1991;96:20721–20733. <http://dx.doi.org/10.1029/91JD02040>
- Sillman S. New developments in understanding the relation between ozone, NO<sub>x</sub> and hydrocarbons in urban atmosphere. *Adv S Phys Chem.* 1995;3:145–171. [http://dx.doi.org/10.1142/9789812831712\\_0005](http://dx.doi.org/10.1142/9789812831712_0005)
- Young AT, Betterton EA, Salazar de Rueda L. Photochemical box model for Mexico City. *Atmosfera.* 1997;10:161–178. Available from: <http://www.redalyc.org/pdf/565/56510401.pdf>
- Kono H, Ito S. A micro-scale dispersion model for motor vehicle exhaust gas in urban areas – OMG Volume-Source model. *Atmos Environ.* 1990;24B:243–251. [http://dx.doi.org/10.1016/0957-1272\(90\)90029-T](http://dx.doi.org/10.1016/0957-1272(90)90029-T)
- Yamartino RJ, Wiegand G. Development and evaluation of simple models for the flow, turbulence and pollutant concentration fields within an urban street canyon. *Atmos Environ.* 1986;20:2137–2156. [http://dx.doi.org/10.1016/0004-6981\(86\)90307-0](http://dx.doi.org/10.1016/0004-6981(86)90307-0)

13. Gifford FA, Hanna SR. Modelling urban air pollution. *Atmos Environ.* 1973;7:131–136. [http://dx.doi.org/10.1016/0004-6981\(73\)90202-3](http://dx.doi.org/10.1016/0004-6981(73)90202-3)
14. Robeson SM, Steyn DG. Evaluation and comparison of statistical forecast models for daily maximum ozone concentrations. *Atmos Environ.* 1990;24B:303–312. [http://dx.doi.org/10.1016/0957-1272\(90\)90036-T](http://dx.doi.org/10.1016/0957-1272(90)90036-T)
15. Jöckel P, Sander R, Kerker G, Tost H, Lelieveld J, Jöckel P. Technical note: The modular earth submodel system (MESSy): A new approach towards earth system modeling. *Atmos Chem Phys.* 2005;5:433–444. <http://dx.doi.org/10.5194/acp-5-433-2005>
16. Butler TM. Automated sequence analysis of atmospheric oxidation pathways: SEQUENCE version 1.0. *Geosci Model Dev.* 2009;2:145–152. <http://dx.doi.org/10.5194/gmd-2-145-2009>
17. Bloss C, Wagner V, Jenkin ME, Volkamer R, Bloss WJ, Lee LD, et al. Development of a detailed chemical mechanism (MCMv3.1) for the atmospheric oxidation of aromatic hydrocarbons. *Atmos Chem Phys.* 2005;5:641–664. <http://dx.doi.org/10.5194/acp-5-641-2005>
18. Jenkin ME, Saunders SM, Wagner V, Pilling MJ. Protocol for the development of the master chemical mechanism, MCM v3 (Part B): Tropospheric degradation of aromatic volatile organic compounds. *Atmos Chem Phys.* 2003;3:181–193. <http://dx.doi.org/10.5194/acp-3-181-2003>
19. Butler TM, Taraborrelli D, Brühl C, Fischer H, Harder H, Martinez M, et al. Improved simulation of isoprene oxidation chemistry with the ECHAM5/MESSy chemistry-climate model: Lessons from the GABRIEL airborne field campaign. 2008;8:4529–4546. <http://dx.doi.org/10.5194/acp-8-4529-2008>
20. Regelin E, Harder H, Martinez M, Kubistin D, Tatum Ernest C, Bozem H, et al. HOx measurements in the summertime upper troposphere over Europe: A comparison of observations to a box model and a 3-D model. *Atmos Chem Phys.* 2013;13:10703–10720. <http://dx.doi.org/10.5194/acp-13-10703-2013>
21. Stickler A, Fischer H, Bozem H, Gurk C, Schiller C, Martinez-Harder M, et al. Chemistry, transport and dry deposition of trace gases in the boundary layer over the tropical Atlantic Ocean and the Guyanas during the GABRIEL field campaign. *Atmos Chem Phys.* 2007;7:3933–3956. <http://dx.doi.org/10.5194/acp-7-3933-2007>
22. Scorgie Y, Annegarn H, Burger L. Socio-economic impact of air pollution reduction measures – Task 1: Definition of air pollutants associated with combustion processes report. Johannesburg: National Economic Development and Labour Council; 2004.
23. Escience Associates Consulted (Pty) Ltd. The Medium term review of the 2009 Vaal Triangle airshed priority area air quality management plan. Draft review report. Pretoria: Department of Environmental Affairs; 2013.
24. Liebenberg-Enslin H, Hurt Q. Gauteng Province Air Quality Management Plan – Final report. Report No. APP/05/CTMM-02a Rev 3. Johannesburg: Gauteng Department of Agriculture and Rural Development; 2009. Available from: <http://www.gdard.gpg.gov.za/Services1/Air%20Quality%20Management%20Plan.pdf>
25. Jaars K, Beukes JP, Van Zyl PG, Venter AD, Josipovic M, Pienaar JJ, et al. Ambient aromatic hydrocarbon measurements at Welgedund, South Africa. *Atmos Chem Phys.* 2014;14:7075–7089. <http://dx.doi.org/10.5194/acp-14-7075-2014>
26. Derwent R, Jenkin M, Saunders S. Photochemical ozone creation potentials for a large number of reactive hydrocarbons under European conditions. *Atmos Environ.* 1996;30:181–199. [http://dx.doi.org/10.1016/1352-2310\(95\)00303-G](http://dx.doi.org/10.1016/1352-2310(95)00303-G)
27. Jorquera H. Air quality at Santiago, Chile: A box modeling approach I. Carbon monoxide, nitrogen oxides and sulfur dioxide. *Atmos Environ.* 2002;36:315–330. [http://dx.doi.org/10.1016/S1352-2310\(01\)00417-4](http://dx.doi.org/10.1016/S1352-2310(01)00417-4)
28. Liebenberg-Enslin H. A review of existing information on air quality issues related to vehicle emissions in South Africa. Report No.: APP/08/SAPIA Rev 2. Johannesburg: South African Petroleum Industry Association; 2008.
29. Goyns P. Modelling real world driving, fuel consumption and emissions of passenger vehicles: A case study in Johannesburg. Johannesburg: University of Johannesburg; 2008.
30. Menut L, Goussebaile A, Bessagnet B, Khvorostyanov D, Ung A. Impact of realistic hourly emissions profiles on air pollutants concentrations modelled with CHIMERE. *Atmos Environ.* 2012;49:233–244. <http://dx.doi.org/10.1016/j.atmosenv.2011.11.057>
31. European Environment Agency (EEA). EMEP/EEA air pollutant emission inventory guidebook. Copenhagen: EEA; 2009.
32. Blasting TJ. Recent greenhouse gas concentrations. Washington, DC: US Department of Energy, Office of Science; 2013.
33. Gonzalez Abad G, Allen NDC, Bernath PF, Boone CD, Mcleod SD, Manney GL, et al. Ethane, ethyne and carbon monoxide concentrations in the upper troposphere and lower stratosphere from ACE and GEOS-ChemPa comparison study. *Atmos Chem Phys.* 2001;1:9927–9941. Available from: [www.atmos-chem-phys.net/11/9927/2011/](http://www.atmos-chem-phys.net/11/9927/2011/)
34. Diab RD, Raghunandan A, Thompson AM, Thouret V. Classification of tropospheric ozone profile over Johannesburg based on mozaic aircraft data. *Atmos Chem Phys.* 2003;3:713–723. <http://dx.doi.org/10.5194/acp-3-713-2003>
35. Rudolph J. The tropospheric distribution and budget of ethane. *J Geophys Res.* 1995;100:11369–11381. <http://dx.doi.org/10.1029/95JD00693>
36. Logan JA, Prather MJ, Wofsy SC, Mcleod MB. Tropospheric chemistry: A global perspective. *J Geophys Res.* 1981;86:7210–7254. <http://dx.doi.org/10.1029/JC086iC08p07210>
37. Salby ML. Fundamentals of atmospheric physics. Cambridge: Cambridge University Press; 1996. [http://dx.doi.org/10.1016/S0074-6142\(96\)80037-4](http://dx.doi.org/10.1016/S0074-6142(96)80037-4)
38. Thompson AM, Diab RD, Bodeker GE, Zuncel M, Coetzee GJR, Arher CB, et al. Ozone over southern Africa during SAFARI-92/TRACE A. *J Geophys Res.* 1996;D19(101):23793–23807. <http://dx.doi.org/10.1029/95JD02459>
39. Thompson AM, Balashov NV, Witte JC, Coetzee JGR, Thouret V, Posny F. Tropospheric ozone increases over the southern Africa region: Bellwether for rapid growth in Southern Hemisphere pollution? *Atmos Chem Phys.* 2014;14:9855–9869. <http://dx.doi.org/10.5194/acp-14-9855-2014>
40. Korhenon K, Giannakaki E, Mielonen T, Fuller A, Laakso L, Vakkari V, et al. Atmospheric boundary layer top height in South Africa: Measurements with lidar and radiosonde compared to three atmospheric models. *Atmos Chem Phys.* 2014;14:4263–4278. Available from: [www.atmos-chem-phys.net/14/4263/2014](http://www.atmos-chem-phys.net/14/4263/2014)
41. Collett KS, Piketh SJ, Ross KE. An assessment of the atmospheric nitrogen budget on the Highveld. *S Afr J Sci.* 2010;106(5/6), Art. #220, 9 pages. <http://dx.doi.org/10.4102/sajs.v106i5/6.220>
42. Hewitt CN. The atmospheric chemistry of sulphur and nitrogen in power station plumes. *Atmos Environ.* 2001;35:387–393. [http://dx.doi.org/10.1016/S1352-2310\(00\)00463-5](http://dx.doi.org/10.1016/S1352-2310(00)00463-5)
43. Venter AD, Vakkari V, Beukes JP, Van Zyl PG, Laakso H, Mabaso D. An air quality assessment in the industrialized western Bushveld Igneous Complex, South Africa. *S Afr J Sci.* 2012;108(9/10), Art. #1059, 10 pages. <http://dx.doi.org/10.4102/sajs.v108i9/10.1059>
44. Sillman S. The relation between ozone, NOx and hydrocarbons in urban and polluted rural environments. *Atmos Environ.* 1999;33:1821–1845. [http://dx.doi.org/10.1016/S1352-2310\(98\)00345-8](http://dx.doi.org/10.1016/S1352-2310(98)00345-8)
45. Stockwell WP, Lawson CV, Saunders E, Goliff WS. A Review of tropospheric atmospheric chemistry and gas-phase chemical mechanisms for air quality modelling. *Atmosphere.* 2012;3:1–32. Available from: <http://www.mdpi.com/2073-4433/3/1/1>
46. Seinfeld J, Pandis S. Atmospheric chemistry and physics: From air pollution to climate change. 2nd ed. New York: Wiley; 2006.
47. Department of Energy. Petroleum Products Act, 1977: Discussion document on the review of fuel specifications and standards for South Africa. Pretoria: Department of Energy; 2011.
48. Amann M, Anderson R, Ashmore M, Depledge M, Derwent D, Grennfelt P, et al. Ground-level ozone in the 21st century: Future trends, impacts and policy implications. London: The Royal Society, 2008. Available from: [http://www.accent-network.org/accent\\_documents/ozone%20report%20web%20pdf%20final.pdf](http://www.accent-network.org/accent_documents/ozone%20report%20web%20pdf%20final.pdf)





# Evidence for aeolian origins of *heuweltjies* from buried gravel layers

## AUTHORS:

Michael D. Cramer<sup>1</sup>  
Johanna von Holdt<sup>2</sup>  
Lesego Khomo<sup>1</sup>  
Jeremy J. Midgley<sup>1</sup>

## AFFILIATIONS:

<sup>1</sup>Department of Biological Sciences, University of Cape Town, Cape Town, South Africa

<sup>2</sup>Department of Environmental and Geographical Science, University of Cape Town, Cape Town, South Africa

## CORRESPONDENCE TO:

Michael Cramer

## EMAIL:

michael.cramer@uct.ac.za

## POSTAL ADDRESS:

Department of Biological Sciences, University of Cape Town, Private Bag X3, Rondebosch 7701, South Africa

## DATES:

Received: 22 Jan. 2015

Revised: 24 June 2015

Accepted: 01 Aug. 2015

## KEYWORDS:

mima-mounds; termite; erosion; vegetation pattern; calcrete

## HOW TO CITE:

Cramer MD, Von Holdt J, Khomo L, Midgley JJ. Evidence for aeolian origins of *heuweltjies* from buried gravel layers. *S Afr J Sci.* 2016;112(1/2), Art. #2015-0025, 10 pages. <http://dx.doi.org/10.17159/sajs.2016/20150025>

Although *heuweltjies* (19–32 m diameter) dominate the surface of much of the southwestern Cape of South Africa, their origins, distribution and age remain controversial. Current hypotheses are that the *heuweltjies* are (1) constructed by the excavation and mounding habits of burrowing animals; (2) the result of erosion by water of areas between patches protected from fluvial action by denser vegetation or (3) the product of localised aeolian sediment accumulation beneath denser vegetation associated with termitaria. At a site where quartz-containing gravels occur on the soil surface in areas between *heuweltjies*, these gravels were found to extend as a relatively intact layer of uniform concentration from the inter-mound area into the mound at the same plane as the surrounding soil surface. This buried layer suggests that *heuweltjies* were either built-up by deposition on a previous soil surface layer or eroded from sediment accumulated above the buried gravel layer. Mounds contain a relatively large proportion of silt consistent with sediment deposition. Mound sediment elemental composition was strongly correlated with that of local shale, indicating a local source of sediment. Pedogenesis was considerably more advanced off- than on-mound. There was no evidence of extensive regional aeolian sediment mantling over the vast area in which the *heuweltjies* occur. These findings and observations support the aeolian deposition hypothesis of *heuweltjie* origins combined with a degree of erosion, rather than a termite bioturbation hypothesis or a predominantly erosion-based hypothesis.

## Introduction

To date three classes of hypotheses have been invoked to explain the regular (i.e. over-dispersed) patterning and the raised circular forms of *heuweltjies*, mounds of about 20 m in diameter that occupy 1.2% of the southwestern Cape landscape.<sup>1</sup> The *bioturbation hypothesis* suggested that they were constructed by earth-moving activities of fossorial mammals and/or termites,<sup>2–4</sup> although the role of mammals has largely been considered to be secondary to termites.<sup>4</sup> Inter-mound distances are considered to be within a range that competing aggressive termites may realistically space themselves.<sup>4</sup> Cramer et al.<sup>5</sup> argued that mounds are far too large to have been made by any known termite. Instead of *heuweltjies* being constructional features, they argued that erosion by water of areas not protected by regularly spaced vegetation clumps is the mechanism responsible for the mounds and their spacing, similar to mounds in Brazil.<sup>6</sup> McAuliffe et al.<sup>7</sup> rejected this *erosion hypothesis* and suggested an *aeolian accretion hypothesis* in which aeolian sediment has accumulated around regularly spaced loci, similar to the formation of coppice dunes (also known as nekhas or nabkhas).<sup>8</sup> McAuliffe et al.<sup>7</sup> argued that harvester termite mounds initiated the regular vegetation patterning by providing nutrient-rich soil loci for plants. Termites are also thought to have provided a local accumulation of calcium oxalate-rich plant biomass, which has facilitated the development of the calcified layer characteristic of many *heuweltjies*.<sup>2,4,7–11</sup>

McAuliffe et al.<sup>7</sup> studied *heuweltjies* on the low-relief plain near the west coast of South Africa with widespread aeolian sand deposits. This voluminous sand supply was implicated as the sediment source for the rather large (about 30 m diameter, 2 m height) *heuweltjies* that commonly occur in that area. These authors used the greater degree of pedological development of off-mound soils to argue for aeolian deposition of sediments on mounds than intervening areas. Despite some differences in on- and off-mound sediment particle sizes,<sup>7</sup> the particle size distributions are broadly similar,<sup>3</sup> suggesting that the sediments on- and off-mounds were originally derived from the same source.

In addition to textural differences between on- and off-mound soils, there are also chemical differences that relate to the process that lead to mound formation. It has been commonly observed that *heuweltjies* have petrocalcic (i.e. calcrete) horizons that are absent in the off-mound areas.<sup>11</sup> McAuliffe et al.<sup>7</sup> concluded that the variability in the development of petrocalcic horizons may be associated with the age of the mounds, with older *heuweltjies* forming more extensive petrocalcic horizons. *Heuweltjie* calcretes are ancient<sup>11</sup> and the source of their Ca is contentious. McAuliffe et al.<sup>7</sup> and others have suggested that source of the Ca was the termites *Microhodotermes viator*, that have been proposed to partially forage off-mound and deposit Ca-rich frass on-mound.<sup>7,11,12</sup> *Heuweltjie* soils also contain higher concentrations of N, P, K, B, Cu, Fe, Mn and Zn,<sup>13,14</sup> and bioaccumulation by termites has also been invoked to explain the accumulation of these other nutrients.<sup>7</sup> Based on <sup>87</sup>Sr/<sup>86</sup>Sr ratios of calcrete that are distinct from underlying rocks, it has, however, been suggested that the source is of marine origin delivered via atmospheric transport.<sup>12</sup> Finally, Ca accumulation on-mound was also suggested to be the consequence of plant nutrient mass-flow concentrating nutrients around plant roots,<sup>5</sup> analogous to the formation of caliche associated with 'islands of fertility'.<sup>15,16</sup> Clearly, the formation of *heuweltjies* and associated calcrete and nutrient-rich soils continues to be contentious.

Identifying previous soil surface levels would provide a test of all three hypotheses above. The *aeolian deposition hypothesis* would predict that surface gravels, indicative of the level of a previous soil surface within a mound, should largely be on the same level as those off-mound. This would indicate that the mound has been constructed on top of a previous soil surface. In contrast, the *bioturbation hypothesis* suggests that termites capable of moving particles less than 1 mm in diameter<sup>17</sup> would cause coarse surface gravels in mounds to sink below those in the off-mound soil.<sup>1</sup> This would occur as termites gradually excavate their subterranean tunnels and storage sites

and as they carry finer soil particles upwards to construct their indurate mounds. Some of these gravels may later be returned to the surface through mammalian bioturbation, for instance by moles, porcupines and aardvarks. If prior to mound formation the land surface contained distributed gravel, the *erosion hypothesis* would predict that surface gravel off-mound would be lower than that within the mound, as off-mound erosion would cause them to sink relative to those protected from erosion within the mound.

We selected a study site in the southwestern Cape in South Africa at which a continuous gravel layer exists in the off-mound surface layer. Relative to the age of *heuweltjies*, this layer may be ancient, as evidenced by associated endemic dwarf succulent plant species and specific plant communities.<sup>18</sup> Also, this layer was found within *heuweltjies* indicating it predates the formation of *heuweltjies* (This gravel layer will be discussed later in the article). We used this gravel layer to test the above three hypotheses, and also used particle size and elemental analysis to provide additional information to substantiate our conclusion that vegetation accelerated aeolian deposition likely accounts for mound formation.

## Methods and Materials

### Study site

We sampled mounds near Worcester (33° 37.416S, 19° 32.226E) in vegetation identified as Robertson Karoo. This site is underlain by Bokkeveld Group metasediments, including shale and quartz veins. The *heuweltjies* that were sampled are situated on the concave lower slopes of a hill (240 m above study site). The site has a mean annual temperature of 16.7 °C and 455 mm annual precipitation, 205 mm of which falls in the coldest quarter and 43 mm in the warmest quarter (<http://www.worldclim.org/bioclim>). The vegetation of the region is low stature (<2 m) and shrubby with many succulents occurring on the mounds. The mound boundaries were readily identifiable by the change in vegetation, with important on-mound genera being *Aloe*, *Euphorbia*, *Cotyledon*, *Tylecodon*, *Searsia* and *Euclea*. The study site was selected because it has gravel on the inter-mound surface (Figure 1a). This gravel layer consists predominantly of quartz, but shale fragments are also present. Quartz veins are evident in the landscape immediately upslope of the *heuweltjies*. Although some *heuweltjies* at the locale contain well-developed layers of calcrete within them, several of the mounds examined in this investigation lacked such layers (Figure 1b).



**Figure 1:** (a) A shallow pit off-mound with surface gravel including quartz that was used to indicate whether the mound was the product of erosion or accretion; (b) a soil pit in a mound with coarse gravel exposed in a side wall and a gravel accumulation visible above an indurate duripan at the base of the pit.

### Mound survey

Seven mounds were selected on terrain with a moderate slope (<10%). For each mound, the perimeter of the mound was defined on the basis of the change in elevation and vegetation characteristics. The down-slope

and across-slope directions were determined and then a Leica NA720 Automatic Level (Leica Geosystems, Johannesburg, South Africa) was used to measure the elevation of the mounds at about 1 m intervals starting from 2 m into the off-mound area downslope and continuing across the mound to 2 m off-mound upslope. The measurements were used to calculate the maximum height of the mound above a linear interpolation between the off-mound points in two orthogonal directions, and these heights were averaged. This allowed estimation of the depth the gravel layer would be expected within the mound. The mound diameter was used to estimate the mound area for the approximately circular mounds. The mound volume was estimated from the area and the average heights of the mounds in the two measured orthogonal directions.

### Gravel, soil and rock sampling

Sampling was carried out from the surface off- and on-mound to the depth of a hard, gravelly duripan. In the case of the biggest mound, this duripan coincided with a strongly cemented petrocalcic horizon. Initial sampling ( $n=7$ ) of mounds was conducted using a 0.07 m diameter auger allowing sampling in 0.22 m increments and ready identification of the location of the gravel layer. Considering the difficulty of getting the auger through the duripan horizons below the gravel layer, we also excavated pits in a subset of mounds ( $n=5$ ). In addition to these mounds, one of the largest mounds in the area was also excavated. For pits, the soil area was marked out in a 0.2 m x 0.2 m area and excavated in 0.2 m intervals. Off-mound sampling was restricted to a single 0–0.2 m depth, but on-mound sampling was to a maximum depth of 0.8 m, apart from the large mound that was sampled to a depth of 1.8 m. The soil excavated by auger or from each 0.2 m x 0.2 m x 0.2 m volume was sieved through a 0.005 m mesh sieve. Care was taken to break up peds so that only coarse gravel remained on the sieve. The coarse gravel and sieved soil were transported to the laboratory and the stones were weighed. The sieved stones were subject to size fractionation using a test sieve shaker (Endecott Ltd, London, UK) and then each fraction was weighed. We were able to get a larger and more representative sample from the pits than from using the auger, so we based measures of gravel concentration, particle size and pH on samples from the pits. The 5 mm sieved soil was air-dried for 48 h and then 1 mm sieved samples were split by repeatedly quartering the soil and taking opposite quarters for each successive sub-sample to provide samples of appropriate volume<sup>19</sup> for particle size analysis and elemental analysis. Samples of shale and quartz from rock outcrops were collected, the surface cleaned with a steel brush followed by washing with a domestic power washer before being crushed.

### Soil particle size analysis

Soil particle size distributions were analysed using a Malvern Mastersizer 2000 (Malvern Instruments Ltd, Malvern, UK) on sieved soils smaller than 1 mm suspended in water and ultrasonically dispersed, stirred and introduced to the laser diffractometer with a Malvern Hydro 2000G wet dispersion unit. The organic material that floated to the surface was removed by hand. Each sample was subjected to 180 s ultrasonic dispersal to ensure complete disaggregation of particles. The proportion of the soil particles in each size class were recorded and plotted. These size classes were then summed into categories representing clay, silt and sand, according to the Wentworth grain size chart.<sup>20</sup>

### Soil and rock elemental and pH analysis

Sieved (<1 mm) soil and crushed rock samples were milled to a fine powder with a mortar and pestle. The powder was placed in sample cups with a polypropylene bottom and analysed in a Spectroscout energy-dispersive X-ray Fluorescence (XRF) analyser (SPECTRO Analytical Instruments, Kleve, Germany). The duration of sampling for each filter was set to 100 s for each of the Pd, Mo and Ta filters and 100 s without a filter. The instrument was calibrated by using a certified standard GBW07312 (National Research Center for CRMs, Beijing, China), for which elemental concentrations were obtained from NOAA Technical memorandum NOS ORCA 68 (1992). The elements measured (all above detection limit in all samples) were Na, Mg, Al, P, K, Ca, Ti, V, Cr, Mn,

Fe, Ni, Cu, Zn, Ga, As, Se, Br, Rb, Sr, Y, Zr, Ba, W, Hg, Tl, Pb, Bi and Th. Soil pH was measured by suspending about 20 g sieved air-dried soil in 50 mL 1 M KCl, stirring with a glass rod, allowing to settle for 1 h and then immersing a pH electrode in the supernatant.

### Mound profile descriptions

Profile descriptions were made in two on- and one off-mound position. Major horizons were identified, measured and sampled (multiple samples were taken in thick horizons). The profiles were excavated to the depth of the impenetrable duripan for one mound and the nearby off-mound sites. For one larger mound, sampling was carried out to a depth of 1.8 m, where a nodular calcrete layer prevented further excavation. Sampling to this depth was only possible because of prior excavation by an armadillo, which had dug through an indurated layer, which appeared to be a combination of duripan and calcrete. These samples were analysed for particle size and XRF elemental analysis as described above. The presence of carbonates was also confirmed by testing the effervescence with the addition of a 10% (v/v) HCl solution. Soil horizon characteristics were described following guidelines for soil description (FAO).<sup>21</sup>

### Statistical analyses

Soil gravel concentration, soil pH and elemental concentrations were analysed using ANOVA with post-hoc Tukey tests and general linear models in R<sup>22</sup> to determine whether there were significant ( $p < 0.05$ ) differences between the values. Linear regression and confidence intervals were calculated using *ggplot2* in R.

## Results

### Mound morphology and profiles

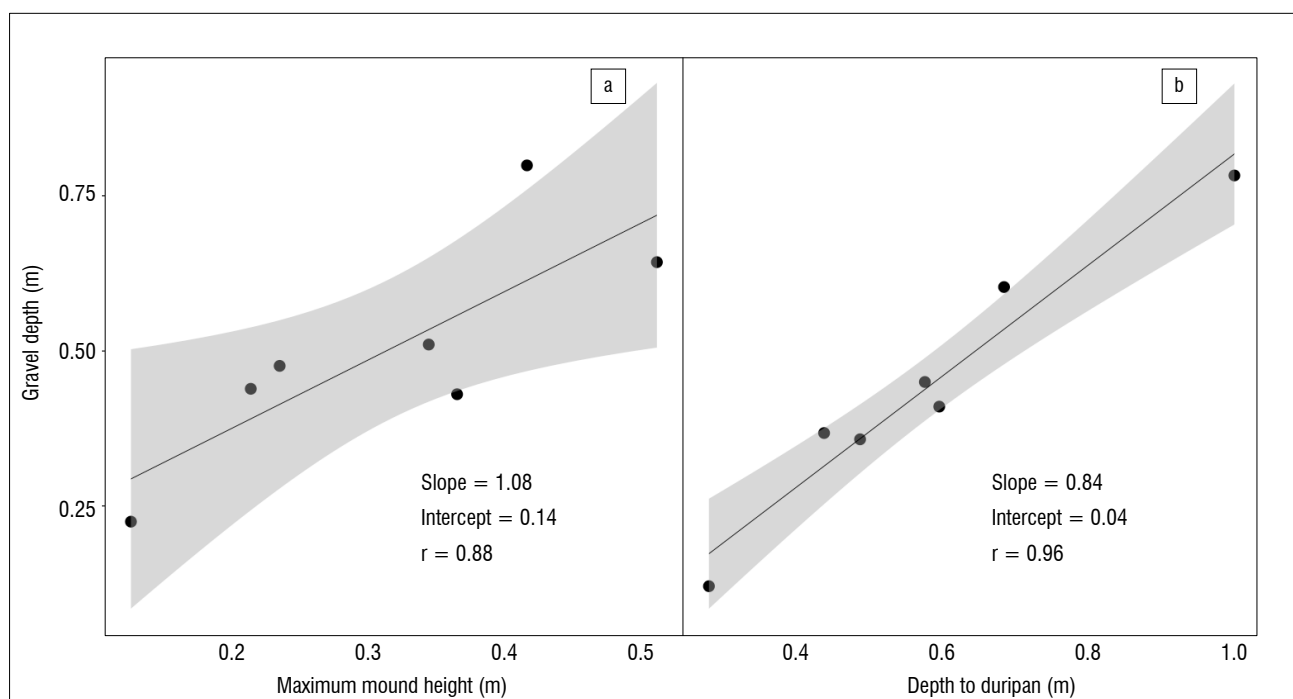
The surveyed mounds had small volumes relative to other mounds in the region (c.f. Cramer et al.<sup>5</sup>) and an average height of only 0.32 m above the landscape surface (Table 1), apart from the one larger mound sampled, which had a height of 1.1 m. Nevertheless, the mounds were clearly demarcated by changes in vegetation and elevation<sup>23</sup> and characterised by a dramatic decrease in the amount of gravel on the soil surface. Excavation by auger revealed that there was a concentration of gravel at 0.48 m below the mound surface on average (Table 1), but

considering the difficulty penetrating below this gravel layer with an auger, we also opened soil pits. In the subset of mounds ( $n=5$ ) in which pits were excavated, the depth at which the maximum concentration of gravel was found with the auger ( $0.54 \pm 0.06$  m) was slightly shallower than from pits ( $0.61 \pm 0.07$  m), but this difference was not significant (Student's paired sample  $t$ -test,  $p=0.121$ ). The difference between the depth at which the gravel layer was found and the expected depth of this layer, based on the maximum mound heights, was not significant (Student's  $t$ -test,  $p=0.076$ ). Mound heights and the depth to the duripan layer were also significantly correlated with the depth at which the gravel layer was found (Figure 2).

**Table 1:** Characteristics of mounds excavated with an auger for which morphological properties were measured ( $n=7$ ).

Mound characteristic	Unit	Mean $\pm$ SE
Maximum landscape slope	%	9.0 $\pm$ 0.6
Area	m <sup>2</sup>	551 $\pm$ 82
Volume	m <sup>3</sup>	140 $\pm$ 41
Maximum mound height above landscape	m	0.32 $\pm$ 0.05
Depth excavated	m	0.57 $\pm$ 0.08
Depth of gravel layer	m	0.48 $\pm$ 0.06
Expected depth of gravel layer	m	0.42 $\pm$ 0.05

In the inter-mound area, the surface gravel was relatively uniformly exposed, or buried below small accumulations of sediment around vegetation (Figure 1a). The off-mound gravel proportions per size class were 0% >31.5 mm, 23% >16 mm, 61% >8 mm and 15% >5 mm, and the on-mound gravel (averaged over all depths) comprised 9% >31.5 mm, 21% >16 mm, 54% >8 mm, 16% >5 mm. The gravel was mainly quartz with some shale fragments. In contrast, the density of gravels on the mound surfaces was much lower, although some gravel was still present throughout the profile. At the depth of the



**Figure 2:** The depth below the mound surface of the maximum accumulation of gravel within each mound (assessed with an auger) was linearly related ( $p < 0.01$ ) to the maximum height of the mounds above the surrounding landscape and to the depth to the duripan layer ( $n=7$ ). The grey band indicates the 95% confidence interval. The slopes, intercepts and Pearson correlation coefficients are shown.

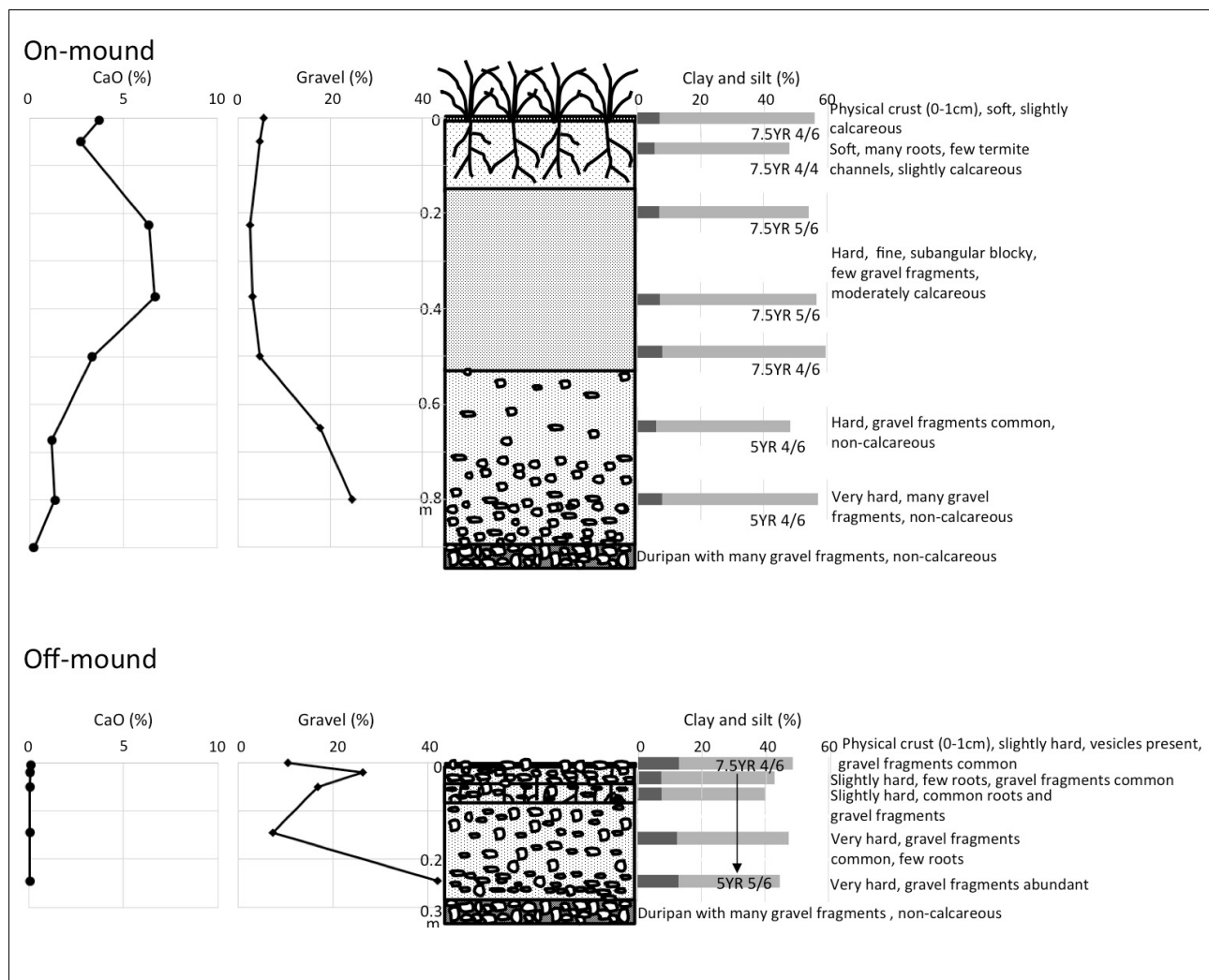
maximum accumulation of gravel within the mound, the average gravel concentration was similar on average to that off-mound, although the variance was high between mounds (Table 2).

The soil profile (Figure 3) confirmed the presence of a concentrated gravel layer situated on top of an impenetrable duripan layer both on- and off-mound. From a sample taken using a hammer and chisel, this gravel

continued into the duripan layer. The profile for this mound showed a significant reduction in gravel in the horizons above this concentrated gravel layer. This mound profile has a horizon that is only moderately calcareous, with no petrocalcic horizons present in the profile. The off-mound profile also has an increased concentration of gravel above an impenetrable duripan layer. Upwards, this gravel concentration first decreases and then increases again towards the surface.

**Table 2:** The variation in the concentration of gravel (kg/m<sup>3</sup>) with elevation in the soil profile above the duripan layer of excavated mounds and corresponding off-mound areas (*n*=5). The cumulative concentration of gravel on-mound is shown for comparison with the off-mound gravel, and the concentrations have also been averaged across mounds. Two of the mounds excavated were less than 0.6 m above the duripan layer in height.

Position	Elevation	Mound					Average
		1	2	3	4	5	
Off-mound	Duripan – 0.2 m	575	430	658	869	210	548
On-mound	0.6–0.8 m	2	27			18	16
	0.4–0.6 m	1	42	48	139	27	51
	0.2–0.4 m	47	612	177	63	22	184
	Duripan – 0.2 m	161	140	942	351	236	366
	Cumulative		211	821	1167	553	303



**Figure 3:** On- and off-mound profiles with CaO (%) of fines, proportion of gravel (% by mass) and proportion of clay and silt of < 1 mm fraction (% by volume). Horizon boundaries are gradual in most instances, but are illustrated here as abrupt for clarity. Gravel fragments and roots are not drawn to scale.



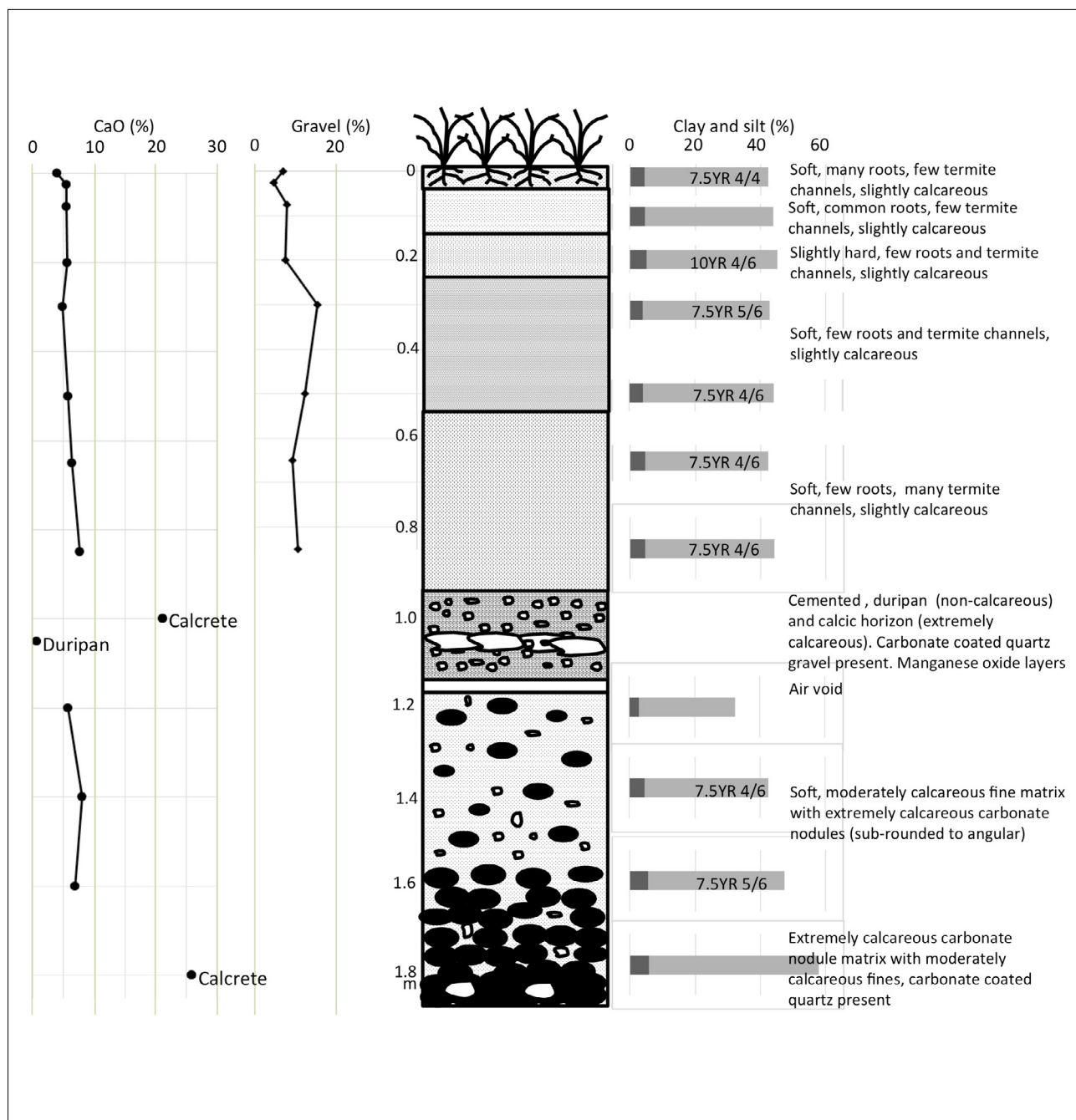
In the large mound for which a profile was described (Figure 4), gravel-sized fragments within a matrix of fines did not occur throughout the profile, although fragments of carbonate nodules did. The diagenetic carbonate accumulations and well-developed petrocalcic horizon occurred at 0.95 m depth and was associated with a duripan layer. This petrocalcic horizon varied from solid calcrete to accumulations of duripan containing no carbonate but also quartz gravel fragments. Below this calcrete is a layer of sub-rounded and angular carbonate nodules ranging in size from pebbles to larger cobbles<sup>20</sup> in a matrix of fines. This nodular carbonate accumulation becomes very dense and excavations were stopped at 1.8 m where it became difficult to penetrate any deeper.

### Soil particle size

Both on- and off-mound soils can be classified as poorly sorted, fine sandy loam at all sampled depths.<sup>24</sup> The lack of sorting is also evident

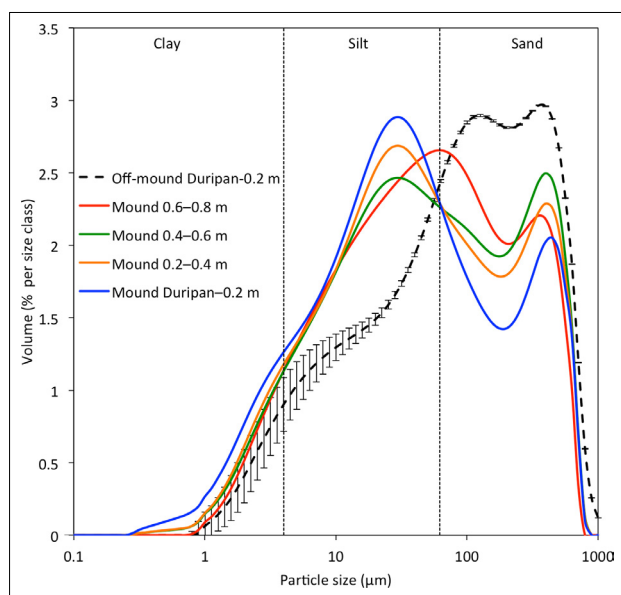
from the particle size distributions (Figure 5), where the off-mound and deepest on-mound horizon (0.6-0.8m) show a clear bimodal distribution. The off-mound soils (0-0.2 m deep) had more sand and correspondingly less silt than the mound soils (Figure 5). The proportions of clay, silt and sand were respectively 7%, 36% and 57% off-mound, and 10%, 50% and 40%, respectively, on-mound (averaged across profile) with silt and sand differing significantly between on- and off-mound ( $p=0.01$ ), but not clay ( $p=0.07$ ). The proportions of clay and silt were greatest closest to the duripan layer within the mounds. The surface mound soil had more very fine and fine sand than the soil from deeper in the mound.

The off-mound profile particle size shows a variation from higher clay content on the surface, decreasing below this and then increasing again towards the duripan (Figure 3). The increased clay content on the surface is because of the vesicular, physical crusting present in the first 0–0.01 m as a result of fluvial deposition of fine material. The average



**Figure 4:** Profile of a large mound with calcrete. CaO (%) of fines is indicated as joined dots and CaO (%) of solid fragments as individual dots. Percentage gravel (% by mass) and proportion of clay and silt of <1 mm fraction (% by volume). Horizon boundaries are gradual in most instances, but illustrated here as abrupt for clarity. Gravel fragments and roots are not drawn to scale.

off-mound particle size distribution (Figure 5) shows a large variation in values for the silt and clay particle range.



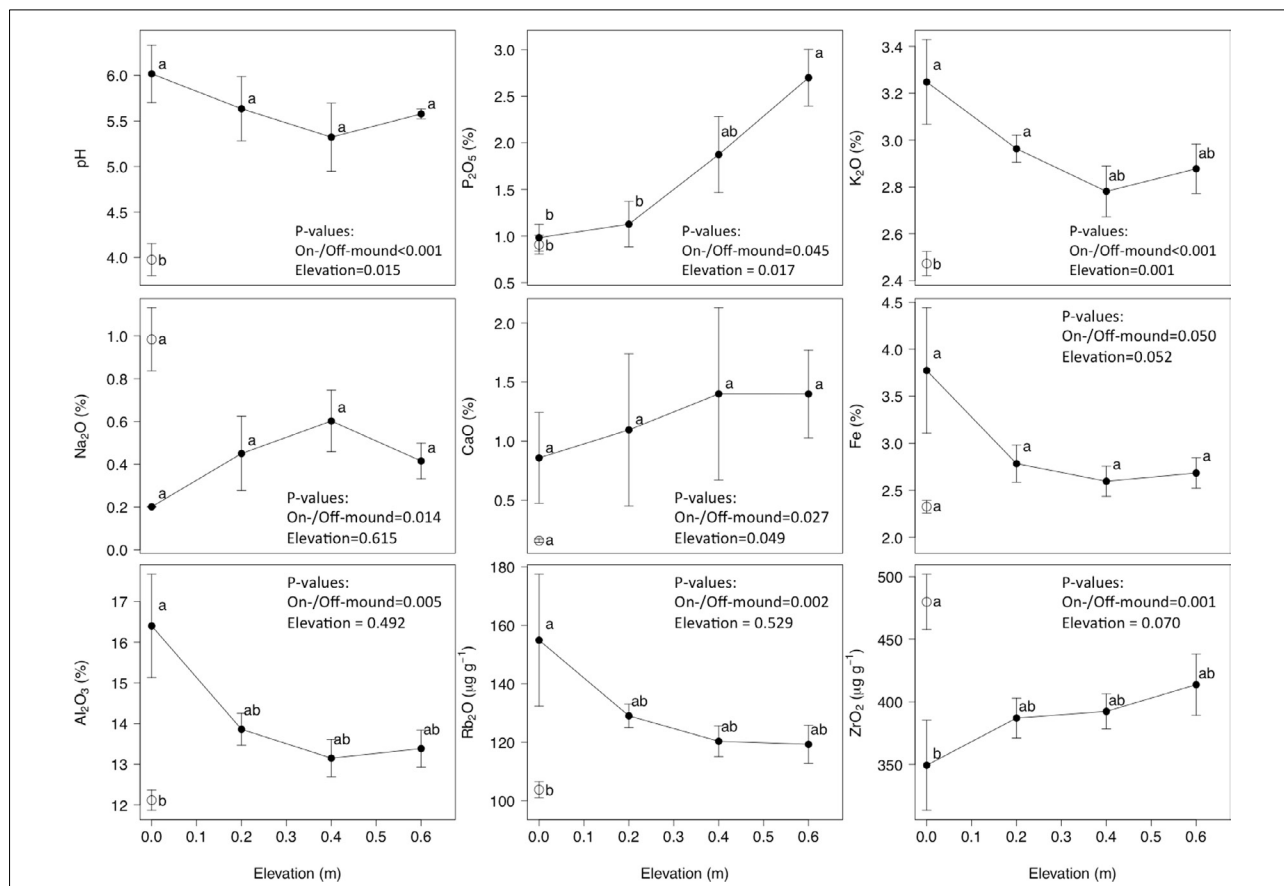
**Figure 5:** The variation in soil particle size between sieved (<1 mm) soil from on- and off-mound sites and with elevation above the duripan layer within mounds. The values represent the average for each particle size fraction ( $n=5$ ). The error bars for the off-mound soil represent standard error. Error bars were not shown for the on-mound for visual clarity.

### Soil and rock chemistry

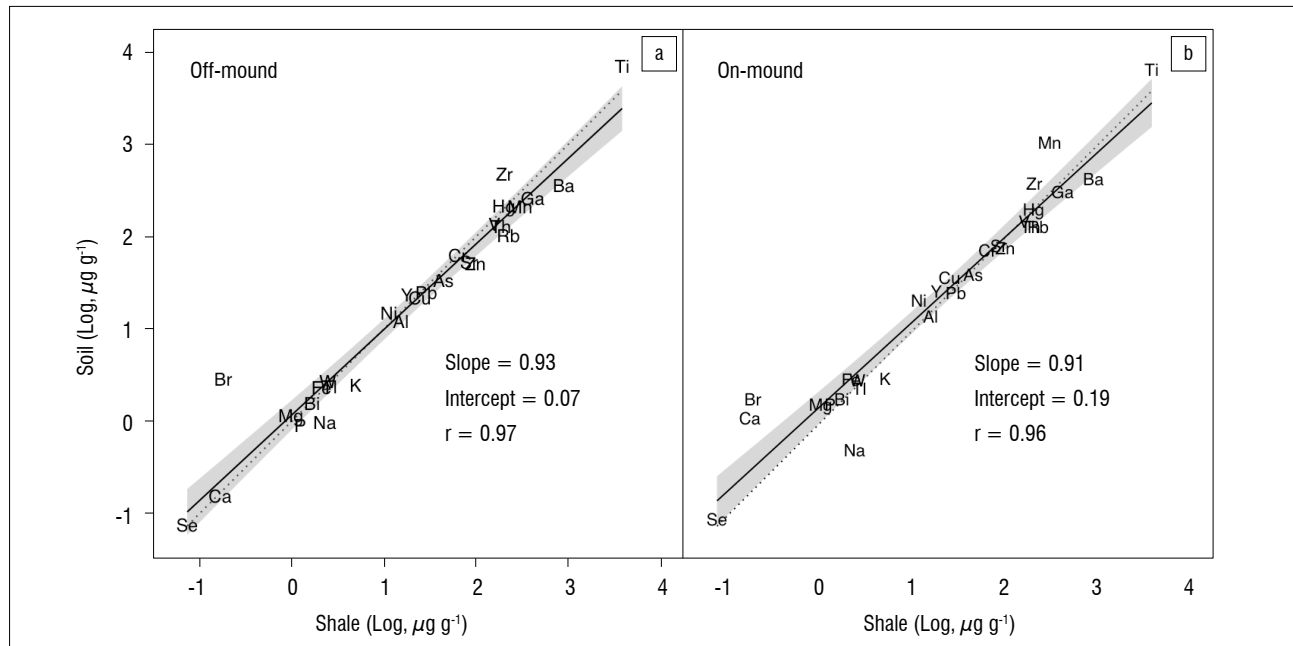
Soil pH was lower off-mound than on-mound and a linear model for pH with depth revealed a significant ( $p=0.015$ ) increase in pH with elevation above the duripan layer on-mound (Figure 6, Table 3). Concentrations of P, K, Ca, Fe, Al and Rb were significantly higher and Na and Zr lower on-mound than off-mound (Figure 6). There were significant increases in P and Ca and decreases in K with elevation above the duripan layer on-mound. Off-mound soil elemental composition was strongly correlated with shale composition (Figure 7), but less strongly related to quartz (i.e. slope=0.89, intercept=0.53,  $r=0.87$ ). Concentrations of Br, Zr and Ti were higher and Na and K lower than the 95% confidence interval on the regression of off-mound soil relative to shale. The elemental composition of on-mound soil (averaged across all depths) was less strongly correlated with shale composition than for off-mound soils (Figure 7). Concentrations of Ca, Br, Zr, Mn and Ti were higher and Na and K lower than the 95% confidence interval in the regression of on-mound soil relative to shale.

### Discussion

We found that all mounds comprised largely fine soils and little gravel with the gravel layer at the same plane as that off-mound. The degree of pedogenesis was greater off- than on-mound, consistent with McAuliffe's<sup>7</sup> findings. Together these observations support the hypothesis that the mound sediments are of aeolian origin, deposited over an existing gravel layer, rather than eroded from a former gravel-containing land-surface. Although traces of termites were found in the mounds, if they had played an important role in *heuweltjie* formation, we would have expected within mound gravels to be substantially lower in elevation than those off-mound.<sup>2</sup> The relative absence of gravel in the upper layers of the mound sediments is also unlikely to be the result of extensive termite bioturbational lowering of gravels, because the gravel layer lower down



**Figure 6.:** The variation of soil pH and elemental concentrations between the on- and off-mound sites and with elevation in the mound profile above the duripan layer. The error bars indicate the standard error ( $n=5$ ) and different letters indicate significant ( $p<0.05$ ) differences as determined by ANOVA followed by post-hoc Tukey tests. The  $p$ -values in each panel are for the general linear model in which site (on- or off-mound) was a categorical predictor and elevation within mound a continuous predictor.



**Figure 7:** The correlation between the shale elemental composition and the elemental composition of on- and off-mound soil. The grey band indicates the 95% confidence interval. The slope, intercept and Pearson correlation coefficient are shown. The element positions are identified by their symbols and include Na, Mg, Al, P, K, Ca, Ti, V, Cr, Mn, Fe, Ni, Cu, Zn, Ga, As, Se, Br, Rb, Sr, Y, Zr, Ba, W, Hg, Ti, Pb, Bi and Th.

**Table 3:** The elemental concentrations of the off- and on-mound soils (averaged across all depths) with shale and quartz rock collected from the study site. The values are the mean ± SE.

Element	Unit	Off-mound (n = 5)	On-mound (n = 5)	Shale (n = 3)	Quartz (n = 3)
Na <sub>2</sub> O	%	0.98 ± 0.15	0.48 ± 0.05	2.29 ± 1.18	0.62 ± 0.05
MgO	%	1.16 ± 0.04	1.56 ± 0.04	0.98 ± 0.21	0.26 ± 0.04
Al <sub>2</sub> O <sub>3</sub>	%	12.12 ± 0.25	14.24 ± 0.7	15.33 ± 0.66	2.57 ± 0.02
P <sub>2</sub> O <sub>5</sub>	%	0.91 ± 0.1	1.47 ± 0.26	1.24 ± 0.37	4.24 ± 4.16
K <sub>2</sub> O	%	2.47 ± 0.05	2.97 ± 0.1	4.92 ± 0.64	0.71 ± 0.05
CaO	%	0.16 ± 0.01	1.06 ± 0.13	0.17 ± 0.04	0.22 ± 0.1
TiO <sub>2</sub>	µg/g	7048 ± 242	6487 ± 264	3858 ± 576	1167 ± 167
V <sub>2</sub> O <sub>5</sub>	µg/g	138 ± 4	148 ± 1	160 ± 17	69 ± 13
Cr <sub>2</sub> O <sub>3</sub>	µg/g	64 ± 3	72 ± 1	64 ± 17	33 ± 2
MnO	µg/g	215 ± 4	1031 ± 109	301 ± 19	91 ± 11
Fe	µg/g	2.3 ± 0.1	2.9 ± 0.2	2.1 ± 0.2	0.4 ± 0.1
CoO	µg/g	8.7 ± 0.2	9.9 ± 0.4	8.4 ± 0.3	3.6 ± 0.2
NiO	µg/g	14.9 ± 0.5	21 ± 0.7	11.4 ± 0.7	4.4 ± 0.7
CuO	µg/g	21.8 ± 1.8	37.5 ± 2	24.6 ± 8.5	6 ± 1.4
ZnO	µg/g	51 ± 0	77 ± 2	99 ± 18	158 ± 28
Ga	µg/g	265 ± 8	318 ± 20	414 ± 35	114 ± 17
As <sub>2</sub> O <sub>3</sub>	µg/g	34 ± 3	41 ± 4	44 ± 13	21 ± 2
Se	µg/g	0.076 ± 0.007	0.089 ± 0.004	0.073 ± 0.007	0.06 ± 0.005
Br	µg/g	2.86 ± 0.52	1.74 ± 0.07	0.18 ± 0.06	0.12 ± 0
Rb <sub>2</sub> O	µg/g	104 ± 3	132 ± 8	225 ± 45	30 ± 8
SrO	µg/g	54 ± 3	80 ± 2	84 ± 27	32 ± 5
Y	µg/g	23.6 ± 1.5	25.7 ± 1.6	17.7 ± 1.1	0.7 ± 0
ZrO <sub>2</sub>	µg/g	480 ± 22	384 ± 9	205 ± 13	16 ± 5
BaO	µg/g	362 ± 19	435 ± 16	899 ± 88	168 ± 57
WO <sub>3</sub>	µg/g	2.7 ± 0.3	2.9 ± 0.2	2.5 ± 0.8	0.9 ± 0.1
Hg	µg/g	218 ± 13	204 ± 14	201 ± 2	212 ± 15
Tl	µg/g	2.4 ± 0.1	2.3 ± 0.1	2.6 ± 0.5	1.5 ± 0.1
Pb	µg/g	25 ± 3.1	25.3 ± 0.9	28.8 ± 8	8 ± 3.5
Bi	µg/g	1.58 ± 0.12	1.78 ± 0.07	1.66 ± 0.1	0.97 ± 0.02
Th	µg/g	129 ± 14	130 ± 4	186 ± 18	9 ± 0

the soil profile is largely intact and at the elevation of the off-mound soil surface. This is substantiated by the fact that the depth of the gravel layer in the mound is correlated with the height of the mound above the surrounding landscape. The sizes of the gravel were also comparable between the on- and the off-mound sites. The proportion of gravel in the smaller mound profile above the concentrated gravel layer was less than 5% of the total amount of gravel in the profile (Figure 3). Thus, although there was some gravel in the mound above the gravel layer, this was a small proportion of the total amount. There was evidence at the study site of significant fossorial activity (rodent, porcupine and aardvark) on the mounds, making it likely that fauna redistributed some of the gravel upwards within the profile of the mounds.

A previous analysis by Cox et al.<sup>2</sup> of the small-sized stone content of *heuweltjies* supported the role of termites and fossorial animals. Because fewer stones were observed on- than off-mound, this led to the erroneous conclusion that bioturbation had led to the sinking of stones through the soil profile. However, these authors noted the limitation of their study and study site; the study mainly focused on surface layers (0–0.15 m depth), it took place in some *heuweltjies* where the concretionary calcrete layer prevented excavation beyond 0.3–0.4 m and these authors did not separate concretionary calcrete rocks from non-concretionary rocks. Furthermore, these authors did not consider that the lack of stones in surface layers may also be the result of aeolian deposition, rather than only being because of bioturbation. Our study site had the advantage of the continuous ancient gravel layer over a duripan, providing an indication of the previous soil surface.

Similar to previous reports,<sup>7,13</sup> there were differences between particles sizes on- and off-mounds. Movement of sediments by wind depends on the sizes of the sediment particles and the power of the wind.<sup>25</sup> Based on a compilation of studies of dust, Lawrence et al.<sup>26</sup> concluded that local dust deposition consists predominantly of coarse silts and/or fine sands, with 10–60% (by mass) of the particles being smaller than 20  $\mu\text{m}$ . Local dust generally also contains large fractions of fine silt (25–60%) and clay (10–40%). Regional dust, on the other hand, consists of less coarse silt and fine sands and a larger proportion of fine silt, but it has a clay content similar to local (within the same geological context) deposition. The particle size analysis reported here with greater concentrations of fine particles on-mound is thus consistent with a local source of sediment, as detailed by Lawrence et al.<sup>26</sup> In contrast, McAuliffe et al.<sup>7</sup> speculated that sediment at their site was derived from the coast and/or nearby dune fields. The local origin of sediments at our site is substantiated by elemental analysis, which showed that both on- and off-mound soils were strongly correlated with the elemental composition of shale rock sampled upslope from the mounds. McAuliffe et al.<sup>7</sup> concluded that sediments at their sites were well sorted. In contrast, the sediment sampled in this study was classified as poorly sorted. This could be the result of the better resolution provided by the laser diffraction particle size analysis compared to the hydrometer method. The occurrence of greater amounts of silt on- than off-mound is consistent with preferential aeolian deposition to the mounds, but accompanied by preferential loss of silt from the off-mound area and formation of a physical crust with vesicular porosity in places because of the wetting and drying cycles of fluvial action.<sup>27</sup>

Strong differences in elemental concentrations between on- and off-mound soils have been noted previously<sup>13,14</sup> and are generally regarded as the consequence of biotically (i.e. termite) induced accumulation of elements on-mound. It is possible, however, that off-mound elemental concentrations may be lower than those on-mound because of greater leaching of off-mound soils as a result of greater drainage through these lower elevation areas. The fact that concentrations of the relatively immobile elements Zr and Ti in off-mound soils are above the 95% confidence band of the correlation between off-mound soil and shale and that the slope of the line is less than 1 indicates that leaching of other more mobile elements has indeed occurred. Despite this general trend, some elements (notably Ca and Mn) have accumulated relative to other elements in mound soils, as the concentrations lie above the 95% confidence band for the correlation between on-mound soils and shale. The fact that Br is high in both on- and off-mound soils relative to shale

is surprising, but may indicate some marine influence.<sup>28,29</sup> Strontium isotopes and  $\delta^{13}\text{C}$  values in *heuweltjie* calcrete at Clanwilliam were not similar to that in surrounding geology, leading to the suggestion that they may be of marine origin.<sup>12</sup> Interception of atmospheric nutrients by vegetation is not unusual<sup>30</sup> and thus, in addition to vegetation trapping aeolian sediment to form the mound and concentrating nutrients, plants may additionally intercept nutrient particulates and aerosols, consequently enriching the soil over prolonged periods.

This site appears to have much in common with the inland mounds studied by McAuliffe et al.<sup>7</sup> Similar to their Soebatsfontein mounds, the mounds in this study showed great diversity in the development of calcic horizons. The biggest mound surveyed had a well-developed calcrete layer (0.95 m) in combination with a reddened (5YR) argillic duripan layer. This calcrete layer appears to have formed in gravelly materials.<sup>31</sup> The smaller mound profiled (Figure 3) was only moderately calcareous at a depth of between 0.2 m and 0.4 m (7.5YR 5/6), but no distinct structural or textural differentiation at this level, and no significant petrocalcic horizons were present in the mound. McAuliffe et al.<sup>7</sup> highlight the fact that the presence or absence of calcic horizons could be the result of significant age or developmental differences between the mounds because of the time-dependence of the formation of such horizons. The two mounds profiled here are in very different positions in the landscape, which could also contribute to significant developmental differences in the mounds. Furthermore, the horizons above the calcrete on the larger mound were softer, less-cohesive and easier to sample than those of the smaller mound, contrary to what one might expect based on the presence of the well-developed calcic horizon. This could imply younger sediments accumulated on an older, previously existing surface. In addition, McAuliffe et al.<sup>7</sup> also identified a 'strongly indurated, silica-cemented duripan' in the off-mound areas at 0.3 m depth. A similar reddened, cemented duripan was present on- and off-mound in the present study area, in conjunction with and just below a concentrated gravel layer. The horizons above this layer on-mound contain similar fine-textured sediments throughout the mound with limited pedogenic alteration, which combined with the lack of gravel supports the role of aeolian deposition.

Of the elements analysed (Figure 7), only Na and Zr were at significantly higher concentrations off- than on-mound (Figure 6). While Na is relatively mobile in soils and could be readily leached from the lower elevation off-mound soils, Zr is considered relatively immobile<sup>34</sup> and is likely to be higher in the off-mound soils because of loss of other elements associated with finer soil fractions. The on-mound accumulation P, K, Ca, Fe, Al and Rb and higher pH relative to off-mound sites may also be associated with the greater silt content of the mound soils, because like these finer particle sizes, these elements increased in concentration with depth through the mound profile. This indicates that there is a degree of pedogenic development within the mounds, substantiating the antiquity of the mounds. The significant increase in P concentration with elevation above the duripan may indicate plant depletion of this regionally scarce resource,<sup>35</sup> particularly from deeper and older horizons within the mound.

The results presented are consistent with aeolian deposition occurring preferentially on mounds as a result of increased surface roughness and sediment trapping provided by vegetation clumps. Alternatively, regional aeolian sediments deposited over a gravel plain might subsequently have been eroded, leaving mounds associated with vegetation clumps that reduce erosion. McAuliffe et al.<sup>7</sup> argued that the absence of pedogenically advanced layers on-mound ruled out the erosion hypothesis, because, if these sediments were part of a much older surface, the on-mound horizons should have undergone substantially more pedogenesis than is evident. Furthermore, for most sites where *heuweltjies* occur there is no evidence of extensive mantling by aeolian sediment, apart from that associated with the mounds. The mounds thus differ from the surroundings by having less pedogenically developed soil and distinct vegetation. Based on texture and elemental analysis, the sediment at our study site and that at the site studied by McAuliffe et al.<sup>7</sup> is of local origin, making it unlikely that it was part of a regional mantling by aeolian sediment.



Although a predominantly erosional hypothesis seems unlikely, fluvial erosion of sediments and channelling of water between mounds is potentially an important process maintaining the sediment and profile differences between on- and off-mound areas. McAuliffe et al.<sup>7</sup> previously noted differences in pedogenic development between surface layers on- and off-mounds and suggested that this reflected the longer period that off-mound soils have had to develop. Our analysis suggests that the buried surface within mounds and associated duripan development does not differ strongly with that off-mound. The existence of physical crust with vesicular porosity and drainage lines at our study site, however, indicates the important contribution of fluvial processes to mound formation, and particularly the off-mound soils. Comparing the *heuweltjie* landscape morphology to analogous modern landforms and processes, like drainage patterns and nebkhas, provides clues to the processes that shaped them. Seifert et al.<sup>8</sup> concluded that the mounds they studied in Arkansas are relict nebkha dunes even though the centimetre-resolution profile of one of the mounds revealed only a slight asymmetry. The absence of features associated exclusively with water erosion, such as distinct terrace structures, potentially precludes an exclusively water erosion hypothesis, especially taking into consideration the moderate slope at the current site. Unlike for mimas in North America,<sup>32</sup> there has never been a suggestion that *heuweltjies* from different parts of South Africa<sup>33</sup> are polygenetic in origin. It is possible, however, that contributory deposition and erosion processes act to different degrees and have site-specific outcomes for pedogenesis. For example, it has been suggested that nebkhas are strongly influenced by local factors such as sediment supply, local climatic conditions, plant density and growth rates and the interaction between these various factors.<sup>36</sup>

## Conclusion

We concur with McAuliffe et al.<sup>7</sup> that *heuweltjies* are predominantly the consequence of aeolian deposition, probably around vegetation clumps. We question the role of termites in the formation of the mounds or their chemical composition. The physical and chemical properties of mounds are consistent with locally derived aeolian sediment deposition, but with some exotic sources of elements such as possibly marine derived Ca and Br delivered via atmospheric transport. We agree that globally, the origins of large earthen mounds are probably polygenetic<sup>32</sup>, and that in other areas, such as the campos de murundus of Brazil<sup>6</sup> and the mimas of Washington State<sup>37</sup>, these may well be primarily formed by erosion, rather than deposition.

## Acknowledgements

We are grateful to Edward Chirwa and Kirsten Packer for assisting with field and laboratory work and are very grateful to the reviewers for their useful comments. Funding for this project was received from the University of Cape Town (URC awards).

## Authors' contributions

All authors participated in conceptualisation and field sampling. M.D.C. and J.v.H. analysed data and wrote the manuscript. J.J.M. and L.K. edited the manuscript.

## References

1. Cramer MD, Midgley JJ. The distribution and spatial patterning of mimas in South Africa suggests genesis through vegetation induced aeolian sediment deposition. *J Arid Environ.* 2015;119:16–26 <http://dx.doi.org/10.1016/j.jaridenv.2015.03.011>
2. Cox GW, Lovegrove BG, Siegfried WR. The small stone content of Mima-like mounds in the South African cape region: Implications for mound origin. *Catena* 1987;14:165–176. [http://dx.doi.org/10.1016/S0341-8162\(87\)80015-2](http://dx.doi.org/10.1016/S0341-8162(87)80015-2)
3. Lovegrove BG. Mima-like mounds (*heuweltjies*) of South Africa: The topographical, ecological and economic impact of burrowing animals. *Symp Zool Soc Lond.* 1991;63:183–198.
4. Moore JM, Picker MD. *Heuweltjies* (earth mounds) in the Clanwilliam District, Cape Province, South Africa: 4000-year-old termite nests. *Oecologia.* 1990;86:424–432. <http://dx.doi.org/10.1007/BF00317612>

5. Cramer MD, Innes SN, Midgley JJ. Hard evidence that *heuweltjie* earth mounds are relictual features produced by differential erosion. *Palaeogeogr Palaeoclimatol Palaeoecol.* 2012;350–352:189–197. <http://dx.doi.org/10.1016/j.palaeo.2012.06.030>
6. Silva LCR, Vail GD, Haidar RF, Sternberg LdaSL. Deciphering earth mound origins in central Brazil. *Plant Soil.* 2010;336:3–14. <http://dx.doi.org/10.1007/s11104-010-0329-y>
7. McAuliffe JR, Hoffman MT, McFadden LD, King MP. Role of aeolian sediment accretion in the formation of *heuweltjie* earth mounds, western South Africa. *Earth Surf Process Landforms.* 2014;39(14):1900–1912. Available from: <http://dx.doi.org/10.1002/esp.3583>
8. Seifert CL, Cox RT, Forman SL, Foti TL, Wasklewicz TA, McColgan AT. Relict nebkhas (pimple mounds) record prolonged late Holocene drought in the forested region of south-central USA. *Quat Res.* 2009;71:329–339. <http://dx.doi.org/10.1016/j.yqres.2009.01.006>
9. Midgley JJ, Harris C, Hesse H, Swift A. *Heuweltjies*; age, origins and vegetation change based on <sup>13</sup>C and <sup>14</sup>C analyses. *S Afr J Science.* 2002;98:202–204.
10. Francis ML, Ellis F, Lambrechts JJN, Poch RM. A micromorphological view through a Namaqualand termitaria (*Heuweltjie*, a Mima-like mound). *Catena.* 2013;100:57–73. <http://dx.doi.org/10.1016/j.catena.2012.08.004>
11. Potts AJ, Midgley JJ, Harris C. Stable isotope and <sup>14</sup>C study of biogenic calcrete in a termite mound, Western Cape, South Africa, and its palaeoenvironmental significance. *Quat Res.* 2009;72:258–264. <http://dx.doi.org/10.1016/j.yqres.2009.04.008>
12. Midgley JJ, Harris C, Harington A, Potts AJ. A geochemical perspective on the origins and consequences of *heuweltjie* formation in the southwestern Cape, South Africa. *S Afr J Geology.* 2012;115(4):577–586. <http://dx.doi.org/10.2113/gssajg.115.4.577>
13. Midgley GF, Musil CF. Substrate effects of zoogenic soil mounds on vegetation composition in the Worcester-Robertson valley, Cape Province. *S Afr J Bot.* 1990;56:158–166.
14. Kunz NS, Hoffman MT, Weber B. Effects of *heuweltjies* and utilization on vegetation patterns in the Succulent Karoo, South Africa. *J Arid Environ.* 2012;87:198–205. <http://dx.doi.org/10.1016/j.jaridenv.2012.05.007>
15. Shreve E, Mallery TD. The relation of caliche to desert plants. *Soil Sci.* 1933;35:99–112. <http://dx.doi.org/10.1097/00010694-193302000-00002>
16. Schlesinger WH, Pilmanis AM. Plant-soil interactions in deserts. *Biogeochemistry.* 1998;42:169–187. <http://dx.doi.org/10.1023/A:1005939924434>
17. Fall S, Brauman A, Chotte J-L. Comparative distribution of organic matter in particle and aggregate size fractions in the mounds of termites with different feeding habits in Senegal: *Cubitermes nikoensis* and *Macrotermes bellicosus*. *Appl Soil Ecol.* 2001;17:131–140. [http://dx.doi.org/10.1016/S0929-1393\(01\)00125-1](http://dx.doi.org/10.1016/S0929-1393(01)00125-1)
18. Schmiedel U, Mucina L. Vegetation of quartz fields in the Little Karoo, Tanqua Karoo and eastern Overberg (Western Cape Province, South Africa). *Phytocoenologia.* 2006;36:1–44. <http://dx.doi.org/10.1127/0340-269X/2006/0036-0001>
19. Gerlach RW, Dobb DE, Raab GA, Nocerino JM. Gy sampling theory in environmental studies. 1. Assessing soil splitting protocols. *J Chemometrics.* 2002;16:321–328. <http://dx.doi.org/10.1002/cem.705>
20. Williams SJ, Arsenault MA, Buczkowski BJ, Reid JA, Flocks JG, Kulp MA, et al. Surficial sediment character of the Louisiana offshore continental shelf region: A GIS compilation. U.S. Geological Survey (USGS) Open-File Report 2006/1195. Reston, VA: USGS; 2007. Available from: <http://pubs.usgs.gov/of/2006/1195/index.htm>
21. Food and Agriculture Organization of the United Nations (FAO). Guidelines for soil description. 4th ed. revised. Rome: FAO; 2006.
22. R Development Core Team. R: A language and environment for statistical computing. Vienna: Foundation for Statistical Computing; 2014.
23. Midgley GF, Musil CF. Substrate effects of zoogenic mounds on vegetation composition in the Worcester-Robertson valley, Cape Province. *S Afr J Bot.* 1990;56:158–166.
24. Blott SJ, Pye K. Gradistat: A grain size distribution and statistics package for the analysis of unconsolidated sediments. *Earth Surf Process Landforms.* 2001;26:1237–1248. <http://dx.doi.org/10.1002/esp.261>

25. Pye K. Aeolian dust and dust deposits. London: Academic Press; 1987. <http://dx.doi.org/10.1016/B978-0-12-568690-7.50010-7>
26. Lawrence CR, Neff JC. The contemporary physical and chemical flux of aeolian dust: A synthesis of direct measurements of dust deposition. *Chem Geol.* 2009;267:46–63. <http://dx.doi.org/10.1016/j.chemgeo.2009.02.005>
27. Figueira H, Stoops G. Application of micromorphometric techniques to the experimental study of vesicular layer formation. *Pedologie.* 1983;33:77–89.
28. Moyers JL, Duce RA. Gaseous and particulate bromine in the marine atmosphere. *J Geophys Res.* 1972;77:5330. <http://dx.doi.org/10.1029/JC077i027p05330>
29. Saiz-Lopez A, Plane JMC. Bromine oxide in the mid-latitude marine boundary layer. *Geophys Res Lett.* 2004;31:4–7. <http://dx.doi.org/10.1029/2003GL018956>
30. Martorell C, Ezcurra E. The narrow-leaf syndrome: A functional and evolutionary approach to the form of fog-harvesting rosette plants. *Oecologia* 2007;151:561–573. <http://dx.doi.org/10.1007/s00442-006-0614-x>
31. Gile LH, Peterson FF, Grossman RB. Morphological and genetic sequences of carbonate accumulation in desert soils. *Soil Sci.* 1966;101:347–360. <http://dx.doi.org/10.1097/00010694-196605000-00001>
32. Johnson DL, Horwath-Burnham J. Introduction: Overview of concepts, definitions, and principles of soil mound studies. *Geol Soc Am. Special Paper.* 2012;490:1–9. [http://dx.doi.org/10.1130/2012.2490\(00\)](http://dx.doi.org/10.1130/2012.2490(00))
33. Cramer MD, Midgley JJ. The distribution and spatial patterning of mima-like mounds in South Africa suggests genesis through vegetation induced aeolian sediment deposition. *J Arid Environ.* 2015;119:16–26. <http://dx.doi.org/10.1016/j.jaridenv.2015.03.011>
34. Kurtz AC, Derry LA, Chadwick OA, Alfano MJ. Refractory element mobility in volcanic soils. *Geology.* 2000;28:683–686. [http://dx.doi.org/10.1130/0091-7613\(2000\)28<683:REMIVS>2.0.CO;2](http://dx.doi.org/10.1130/0091-7613(2000)28<683:REMIVS>2.0.CO;2)
35. Cramer MD, West AG, Power SC, Skelton S, Stock WD. Plant ecophysiological diversity. In: Allsopp N, Colville JF, Verboom GA, Cowling RM, editors. *Fynbos: Ecology, evolution, and conservation of a megadiverse region.* Oxford: Oxford University Press; 2014; p. 248–272. <http://dx.doi.org/10.1093/acprof:oso/9780199679584.003.0011>
36. Tengberg A, Deliang C. A comparative analysis of nebkhas in central Tunisia and northern Burkina Faso. *Geomorphology.* 1998;22:181–192. [http://dx.doi.org/10.1016/S0169-555X\(97\)00068-8](http://dx.doi.org/10.1016/S0169-555X(97)00068-8)
37. Cramer MD, Barger NN. Are mima-like mounds the consequence of long-term stability of vegetation spatial patterning? *Palaeogeogr Palaeoclimatol Palaeoecol.* 2014;409:72–83. <http://dx.doi.org/10.1016/j.palaeo.2014.04.026>



# ***Australopithecus robustus* societies – one-male or multimale?**

**AUTHOR:**

Katarzyna A. Kaszycka<sup>1</sup>

**AFFILIATION:**

<sup>1</sup>Department of Human Evolutionary Ecology, Institute of Anthropology, Adam Mickiewicz University, Poznan, Poland

**CORRESPONDENCE TO:**

Katarzyna A. Kaszycka

**EMAIL:**

kaszycka@amu.edu.pl

**POSTAL ADDRESS:**

Institute of Anthropology,  
Adam Mickiewicz  
University, Umultowska 89,  
Collegium Biologicum,  
Poznan 61-614, Poland

**DATES:**

**Received:** 04 May 2015

**Revised:** 25 July 2015

**Accepted:** 03 Aug. 2015

**KEYWORDS:**

australopithecine; sexing fossils; size dimorphism; mating system; social structure; k-means clustering

**HOW TO CITE:**

Kaszycka KA. *Australopithecus robustus* societies – one-male or multimale? *S Afr J Sci.* 2016;112(1/2), Art. #2015-0165, 8 pages. <http://dx.doi.org/10.17159/sajs.2016/20150165>

Determining the sex of individual specimens is important in estimating the degree of sexual dimorphism. Sexual dimorphism, in turn, provides clues for reconstructing the social organisation and mating systems of extinct species. In an article published in *Science*, Lockwood et al. (Lockwood CA, Menter CG, Moggi-Cecchi J, Keyser AW. Extended male growth in a fossil hominin species. *Science*. 2007;318:1443–1446.) suggested an uneven sex ratio (in favour of males) for the known individuals of the South African Pleistocene hominid, *Australopithecus robustus*, and claimed evidence of an extended period of growth (delayed maturity) for the males of this species. They concluded that this finding, combined with estimates of sexual size dimorphism, suggests a polygynous reproductive strategy, and a social system similar to that of silverback gorillas (i.e. one-male harems). On re-examination of these claims, and based on further analysis, I agree with Lockwood et al. that morphologically *A. robustus* exhibits an increased (almost gorilla-like) level of facial dimorphism, but propose using an alternate (clustering) technique for grouping the specimens of highly dimorphic species into sexes, and argue that their pronouncements regarding a polygynous social structure of these early hominids are inconclusive. I contend instead that the habitat occupied by this species suggests rather that a one-male harem social structure would have been counterproductive.

## **Introduction**

Sexual dimorphism (genetically determined differences between the sexes) is a common phenomenon among both extant and extinct primates.<sup>1</sup> The dimorphism in secondary sexual characteristics (i.e. those not directly related to reproduction) can be found in a variety of morphological features. In primates, these characteristics are primarily differences in overall body size and dimensions of the canine teeth and are among the skeletal features which can also be observed in fossils. Sexual dimorphism correlates with patterns of social organisation and mating systems.<sup>2-5</sup> As a general rule, size sexual dimorphism is lacking among monogamous primate species, while polygynous species (showing intensive male-male competition) are dimorphic,<sup>6</sup> with males commonly being larger than females. There are, however, two ways in which a difference in body size between males and females at adulthood (i.e. sexual maturity) can be attained during development<sup>7,8</sup>: difference in rate of growth – males growing faster than females, and difference in duration of growth – males maturing, breeding and achieving full body weight later than females (so-called bimaturism). For example, female gorillas start breeding at the age of 10, while male gorillas only 5 years later (at about 11 years of age, males start becoming larger than females).<sup>9</sup>

Sexual dimorphism provides clues for reconstructing the social structure and mating systems of extinct species.<sup>4,10</sup> The current data for early hominids – the australopithecines – are not unequivocal in this regard, suggesting a unique (among primates) combination of small canine size dimorphism and marked body mass dimorphism<sup>4</sup> which makes inferring of social behaviour difficult. In addition, there is also disagreement about the degree of body size dimorphism. Depending on the australopithecine species, the literature provides data of assessed sexual dimorphism ranging in values from characteristics of gorillas<sup>4,11</sup> to those of chimpanzees<sup>12,13</sup> or even modern humans.<sup>14,15</sup> (I use the term 'hominid' rather than 'hominin' to refer to the human clade after the split from the chimpanzee line in its classical, narrow meaning – as a common name (such as 'ape') and not in a taxonomic sense. I use the term 'australopithecines' to refer to a group of species of Plio-Pleistocene extinct relatives of humans, which comprise the so-called 'gracile' and 'robust' forms. Because I disagree with the generic separation of the two groups, I have used the species name *Australopithecus robustus* and not *Paranthropus robustus*.)

The lack of agreement about the degree of size skeletal dimorphism occurs not only for different species of early hominids, but also within the same species and even the same samples. For example, Reno<sup>14</sup> claimed that sexual dimorphism for *Australopithecus afarensis* was not pronounced, i.e., gorilla-like (as generally believed), but only moderate, i.e., contemporary human-like, and implied a principally monogamous mating system for this species – conclusions that were both criticised by Plavcan et al.<sup>16</sup> Lee<sup>17</sup>, on the other hand, argued that size sexual dimorphism of *A. afarensis* varies, depending on the skeletal elements taken into consideration, i.e. in femoral variables it is similar to gorillas, in humeral variables, similar to humans, and in canine variables, similar to chimpanzees. Yet, whatever the degree of dimorphism in the postcranial skeleton of the South African 'robust' australopithecines, it at least seems likely that they had markedly dimorphic faces.<sup>18-20</sup>

In 2007, Lockwood et al.<sup>18</sup> published a paper claiming evidence of an extended period of growth (delayed maturity) for males in the South African *Australopithecus robustus* sample from approximately 1.5–2.0 million years ago, along with its implications. They argued that ranking a sample of early hominid cranial remains on the basis of the stages of tooth wear revealed a difference in size between young and old adult males. The authors<sup>18</sup> stated that their finding, combined with estimates of gorilla-like level of sexual size dimorphism (based on dimensions of the face), suggests a reproductive strategy in which males control mating access to multiple females in a manner similar to that of silverback gorillas. Males, however, experience the costs of such behaviour in the form of high rates of predation, i.e. the peripheral, solitary life of young adult males places them at greater risk of becoming victims of a predator attack.

These claims were then highly publicised and commented upon,<sup>21</sup> but warrant re-examination, as the line of reasoning taken (both for the methods used and interpretation of the results) was somewhat speculative and raises several questions. These concern not only sexing individual australopithecine specimens (the actual sex ratio in the sample, the relation between sex and size, and sex and range of variation), but also the supposed evidence for bimaturism and one-male-multifemale social grouping and lastly, the possibility of a polygynous gorilla-like harem pattern of social structure on savannah.

## Re-examination of Lockwood's claims

### 'Sexing' australopithecine specimens

#### Sex ratio in the *A. robustus* sample

Lockwood et al.<sup>18</sup>, on the basis of size and morphology, concluded that the known craniofacial fragments of the South African robust australopithecines represent individuals in an uneven sex ratio of 4:1 in favour of males. According to the authors,<sup>18</sup> the sample included as many as 15 males and only 4 females. (Previous estimates for the Swartkrans plus Kromdraai sample<sup>22,23</sup> were shown to be biased toward females). The approach I used to determine the actual sex ratio in the sample of *A. robustus* involved sorting the individuals into two groups (presumably sexes) using cluster analysis (a *k*-means clustering algorithm).<sup>24</sup>

#### Materials: samples and measurements

The materials used in my analysis consisted mostly of the same fossil sample of Pleistocene hominids as studied by Lockwood et al.<sup>18</sup> – *Australopithecus robustus* from South African sites of the Gauteng area, housed in the Ditsong National Museum of Natural History (formerly known as the Transvaal Museum) in Pretoria, and the University of the Witwatersrand, Johannesburg. I also used comparative samples consisting of extant African apes (see Figure 1) – western gorillas (*Gorilla gorilla*) and chimpanzees (*Pan troglodytes*) of known sex, both from the Powell-Cotton collection in the Quex Museum, Birchington, Kent, England. (These apes were shot in the wild during the 1920s and 1930s in the French Congo, now known as the Democratic Republic of Congo, and Cameroon).

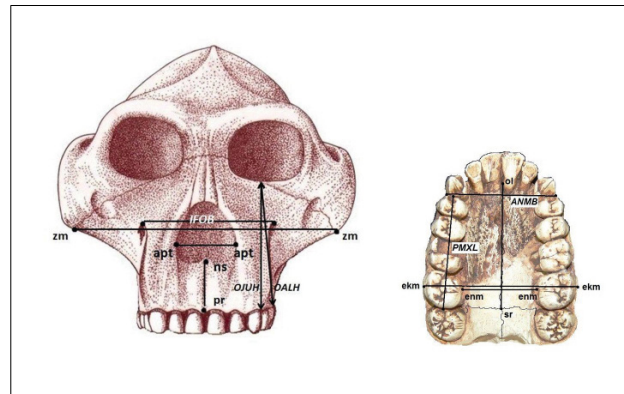


Photo: KA Kaszycka

**Figure 1:** Cranial sexual dimorphism in gorillas (top row) and chimpanzees (bottom row).

Lockwood et al.<sup>18</sup> examined 12 measurements (selected dimensions of the face and palate) of 19 *A. robustus* specimens (17 from Swartkrans, 1

from Kromdraai and 1 from Drimolen). The facial measurements used were: orbito-alveolar height (OALH), orbito-jugal height (OJUH), foraminal height (FORH), alveolar height (ALVH; *ns-pr*), bimaxillary breadth (BMAB; *zm-zm*), interforaminal breadth (IFOB), nasal aperture breadth (NASB; *apt-apt*), and snout breadth (SNOB). The palatal measurements used were: anterior maxillo-alveolar breadth (ANMB), maxillo-alveolar breadth (MAXB; *ekm-ekm*); palatal breadth (PALB; *enm-enm*), and postcanine maxillo-alveolar length (PMXL) (see Figure 2 and Lockwood<sup>19</sup> for measurement definitions).



Drawing on the left: KA Kaszycka<sup>25</sup>

**Figure 2:** Measurements of the face (left) and palate (right) used in the current analysis.

Eleven of these dimensions (excluding snout breadth) were also measured by me. I, however, omitted foraminal height, as I considered this measurement to be not only correlated with the two other height measures of the face (orbito-alveolar height, and orbito-jugal height), but also of little potential use in describing sexual dimorphism. To equate to the same number of dimensions used by Lockwood et al.<sup>18</sup> – i.e., twelve, I added two measurements which were highly dimorphic in apes<sup>20</sup> and of sufficient number in my fossil sample. These two additional dimensions were: maxillary canine (or canine socket) bucco-lingual breadth and palate length, i.e. length of the maxillary palatal process (orale-staurion; *ol-sr*). The above-mentioned bucco-lingual canine dimension also appeared to be quite dimorphic for *A. robustus* (as assessed by the CV method<sup>26</sup> – the index of sexual dimorphism  $ISD=1.10$ ;  $n=26$ ).<sup>20</sup> To make all dimensions comparable, they were standardised for mean values and standard deviations.

My sample of *Australopithecus robustus* included 17 maxillofacial specimens: 14 from Swartkrans Member 1 (dated to ~ 1.6–1.8 Ma) (SK 11, SK 12, SK 13/14, SK 46, SK 48, SK 52, SK 79, SK 83, SK 845, SKW 8, SKW 11, SKW 12, SKW 29, and SKX 265), and one each from Kromdraai (TM 1517), Coopers (COB 101) and Drimolen (DNH 7). Although two specimens included in my analysis – SK 13/14 and SK 52 – were subadults (their third molars were only just erupting), I decided to use them because the dimensions of their canine sockets seemed to sufficiently distinguish them with regard to sex, and their face/palatal measurements were considered to be probably close enough to the minimal values for young adults.

Of the sample of 62 gorilla cranial specimens measured by me,<sup>20</sup> 48 complete (based on possession of at least 10 out of the 11 above-mentioned facial measurements plus the maxillary canine breadth) adult individuals were used for this analysis (26 females and 22 males). In the case of the chimpanzees, 38 out of 56 measured cranial specimens<sup>20</sup> were used (22 females and 16 males). All the apes were adults, based on the eruption of the third molar. All dimensions were measured by the current author, with the exception of the DNH 7 specimen from Drimolen, for which I used Lockwood's<sup>18</sup> and Moggi-Cecchi's<sup>27</sup> data.

#### Methods

My approach to the issue of 'sexing' the australopithecine specimens, and determining the actual sex ratio in the fossil sample of *A. robustus*, was to partition the individuals into two groups (presumably sexes)



using cluster analysis (*STATISTICA 9.1*). The *k*-means clustering technique, through the Hartigan and Wong algorithm, was used<sup>24</sup> (where *k* is the number of clusters desired, while the action in the algorithm centres around finding the *k*-means). Cluster analysis is a multivariate analysis technique that seeks to organise information about variables so that relatively homogeneous groups (clusters) can be formed. From the point of view of the calculation, the *k*-means method is a 'reversal' of analysis of variance (ANOVA). In the *k*-means clustering, the program transfers objects (cases) between *k* randomly formed groups (clusters), to achieve a minimum variation within each group and the maximum variation between groups – so as to form groups (clusters) which will be as different as possible.

The accuracy of the *k*-means clustering was tested on samples of African apes of known sex, which made it possible to compare the results of the cluster analysis with actual data on the sex of the individuals, using the same measurements taken for the australopithecines (i.e. 11 face measurements plus one canine measurement).<sup>20</sup> To assess whether this technique is (or is not) biased towards recovering groups of equal size,<sup>28</sup> and sensitive to the order the algorithm is applied to the cases,<sup>28</sup> several additional analyses were performed using different variants: (1) two, three, and four times as many representatives of one sex than representatives of the other sex in the sample, and (2) different orders of case entry: (a) all representatives of one sex, then all representatives of the other sex, (b) one representative of each sex one after another, and (c) five representatives of one sex, and five representatives of the other sex.

### Results

By applying cluster analytic techniques for the apes, I obtained accurate results for highly dimorphic gorillas, but the method was not as statistically robust for the much less dimorphic chimpanzees. Thus the gorillas were partitioned into two sexes with a high level of accuracy: as many as 94% of the individuals classed among members of the larger and the smaller groups were actually males and females, respectively. Out of 48 individuals only three (6%) were assigned the incorrect sex group: one large gorilla female was grouped with males, and two small gorilla males were grouped with females. Average measurements for actual sexes and the calculated 'sexes' (i.e. those obtained using the clustering algorithm) were insignificantly different from each other. The results for weakly dimorphic chimpanzees, on the other hand, were not accurate, with nearly 30% of individuals (11 out of 38) assigned to the incorrect sex group. There is a significant difference in the frequency of correct sex assignment between gorillas and chimpanzees – according to Fisher's exact test;  $p=0.007$  (Table 1).

**Table 1:** The frequency of correct vs incorrect sex assignment in gorillas and chimpanzees using the *k*-means clustering technique

	Correct sex assignment	Incorrect sex assignment	Total
Gorilla	45	3	48
Chimpanzee	27	11	38
Total	72	14	86

Fisher's exact test: The two-tailed  $p=0.0070$

Having an unequal sex ratio in the gorilla sample: twice as many males than females (22 M vs 11 F), and twice as many females than males (26 F vs. 13 M) had no bearing on the results. Likewise, having three (21 M vs. 7 F), and even four-times (20 M vs. 5 F) as many males than females in the sample also did not significantly change the final classification – each time from zero up to no more than three individuals were assigned the incorrect sex group (0–10%). I therefore reject the hypothesis that the *k*-means clustering algorithm used here is biased towards recovering clusters of roughly equal size. Changing the order of a dataset entry (either the cases or the variables), did not influence

the classification at all. Therefore, I contend that *k*-means clustering is a reliable technique for sorting individual specimens of highly dimorphic species (such as the gorilla) into sexes.

The degree of size sexual dimorphism for *A. robustus* seems to correspond with that for gorillas (the male facial features being on average 17% larger than in females),<sup>18,20</sup> so the *k*-means clustering technique, by assumption, should also be useful in grouping this fossil species into two subsets (sexes). The program divided the *A. robustus* sample into two groups (which differed significantly in 6 out of 12 features<sup>20</sup>) of roughly equal numbers of individuals of both sexes, and not a ratio of 4:1 in favour of the males. In the first group, consisting of the smaller individuals (presumably females), the following specimens were found: SK 13/14, SKW 11, SK 48, DNH 7, COB 101, SK 79, SKW 12 and SKW 8 ( $n=8$ ; Figure 3). In the second group, composed of the larger individuals (presumably males), the following specimens were found: SK 52, SKX 265, TM 1517, SKW 29, SK 845, SK 11, SK 83, SK 46 and SK 12 ( $n=9$ ; Figure 4). One may conclude therefore that either the size sexual dimorphism in *A. robustus* was not as pronounced, and the clustering method used here inadequate, or that the sex ratio reported in the Lockwood et al.<sup>18</sup> study is not necessarily correct.

Other factors used as a 'guide' to variation and sex

### Relation between sex and size

In light of the foregoing, the sex assignment for the specimens of *Australopithecus robustus* in the Lockwood et al.<sup>18</sup> study appears unconvincing. Furthermore, sex assignment can also be questioned from standpoints other than research above. In the Lockwood et al.<sup>18</sup> study, only individuals of overall size categories 1–2 were considered to be females and all others (size categories 3–9) to be males (see Figure 5). In Lockwood's 'Supporting Online Material'<sup>18</sup>, this step was justified in morphological and metric terms, for example, by referring to the smallest (and the youngest) *A. robustus* individual identified as male (SKW 11) and its measurements. These measurements were supposedly found to be closer to that of five other specimens ('males') than they were to the small Drimolen female – DNH 7 – even though the authors<sup>18</sup> recognised that DNH 7 is substantially smaller than other well-preserved skulls of *A. robustus*. Listed in Table 2 are three facial measurements cited by Lockwood et al.<sup>18</sup> to support their statements [see Lockwood's 'Supporting Online Material' p. 5, and Table S1<sup>18</sup>]. In these examples [at least for ANMB and IFOB – see Table 2], it appears that the claim for the measurements of Swartkrans specimen SKW 11 being closer to those of 'other males' than those of the Drimolen female, is an over-interpretation.

**Table 2:** Facial measurements (in mm) of *Australopithecus robustus* individual SKW 11 compared with DNH 7 and other specimens identified by Lockwood et al.<sup>18</sup> as 'males'.

Individual	Measurements		
	ALVH	ANMB	IFOB
DNH 7 (♀)	28	38	45
SKW 11	32	39	48
Range of measurements of 5 other supposed 'males'	30–37	42–51	52–55

ALVH, alveolar height; ANMB, anterior maxillo-alveolar breadth; IFOB, interforaminal breadth

### Relation between sex and range of variation

Lockwood et al.<sup>18</sup> also asserted that 'it is unlikely that any of the other relatively complete skulls are female' (p. 1444), including the smallest of the previously known ones: Kromdraai TM 1517 and Swartkrans SK 48, because the size differences between either of these

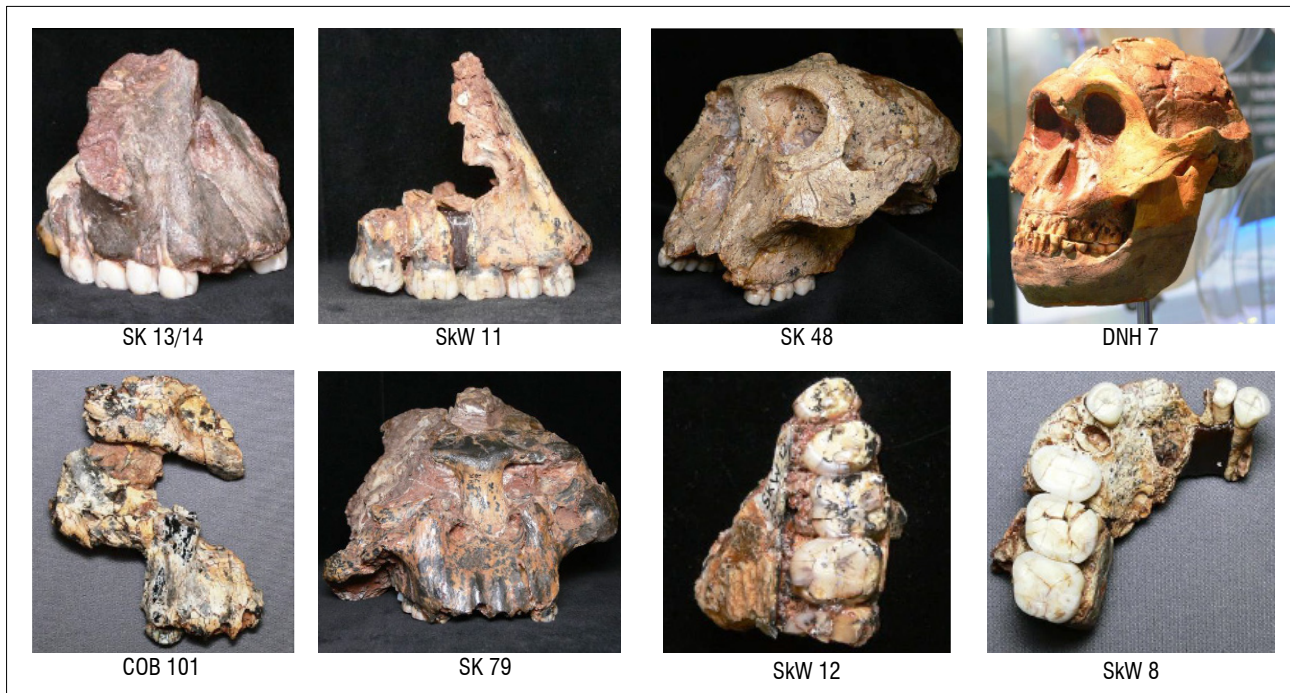


Photo: KA Kaszycka

**Figure 3:** Remains of southern African *Australopithecus robustus* specimens grouped by the *k*-means clustering technique into a cluster of smaller individuals (presumably females).

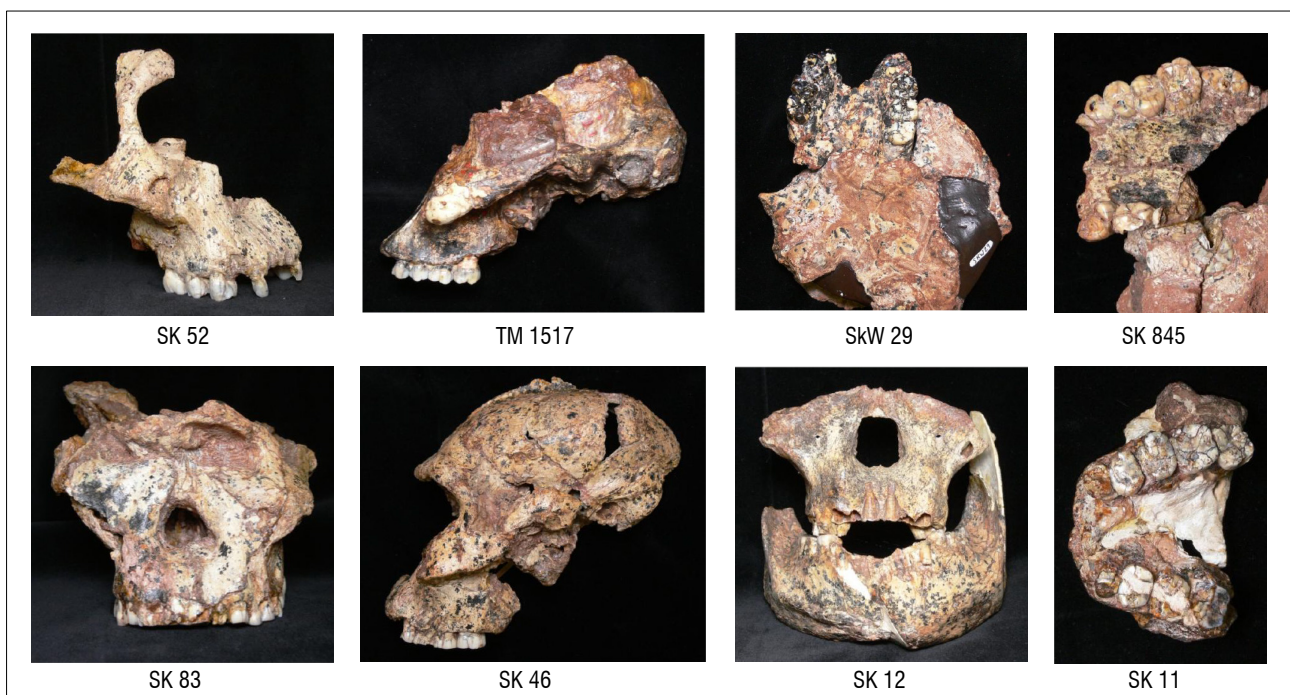


Photo: KA Kaszycka

**Figure 4:** Remains of southern African *Australopithecus robustus* specimens grouped by the *k*-means clustering technique into a cluster of larger individuals (presumably males).

and Drimolen DNH 7 is too large (as is the range of variation among the supposed females).

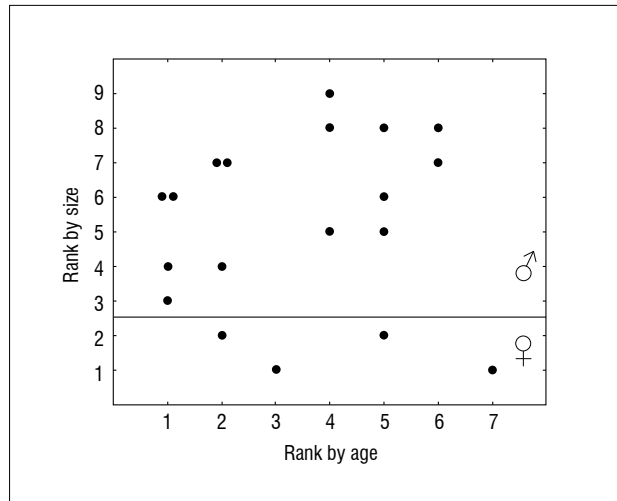
In contrast, firstly, cluster analysis places the individual SK 48 in the same group as DNH 7, therefore these specimens can well be regarded as large and small adult females, respectively. It is worth mentioning that SK 48 was grouped with females in spite of a low sagittal crest. (While crests are

commonly found in gorillas males, gorilla females with crests are known to occur, as was the case for the largest gorilla female in my sample).

Secondly, the range of variation expressed by the ratio of size: 'Mean of a larger individual to the mean of a smaller individual' is greater between the largest and the smallest western gorilla females in a sample  $n=26$ , shot two years apart in Cameroon (ratio=1.15), than between the australopithecine individuals: TM 1517 / DNH 7 (ratio=1.131) or



SK 48 / DNH 7 (ratio=1.128), which may be tens, or even hundreds of thousands years apart. The chance of sampling two gorilla females, whose ratio of size is equal to or greater than 1.13, is  $p=3.4\%$  (probability based on the exact resampling of all 325 possible combinations from a sample of  $n=26$ ). This probability exceeds 2 standard errors ( $SE=1.01$ ) and therefore cannot be discounted. It should also be mentioned that in Lockwood et al.'s Table S2<sup>18</sup>, the given chance of sampling two gorilla females, whose ratio of size is equal to or larger than 1.13, is even greater ( $p=5.4\%$  probability based on 5000 random samples with replacement in a sample  $n=37$ ).



Redrawn: KA Kaszycka from Lockwood et al.<sup>18</sup>

**Figure 5:** Comparison of size ranks to age ranks for adult maxillofacial specimens of *Australopithecus robustus*.

### Bimaturism and social structure

#### Relation between age of adults and size – evidence for bimaturism?

There are several problems with the assessment of correlation between the age of adults and overall size (small size = young adults, large size = old adults) in the *A. robustus* sample, and the claim of evidence of extended male growth. Firstly, when calculating the correlation between age and size of the male specimens, the assumption was made that size categories 3 and 4 are male. In light of the results of the *k*-means clustering analysis I performed, this assumption is not necessarily correct. Furthermore, it is also noteworthy that size-3, age-1 Swartkrans SKW 11 specimen was shown here to be well within the female size range (see Table 2).

Secondly, there may be a problem with assessing the level of statistical significance of the above-mentioned correlation. Lockwood et al. claim that: 'When a randomisation test of correlation coefficient is used, age and size are significantly correlated among the male maxillofacial specimens ( $r=0.52$ ;  $p=0.027$ , one-tailed test)' (p. 1445).<sup>18</sup> Whether or not the one-tailed test should have been used is a matter for debate. The  $p$ -value of the two-tailed test ( $p=0.054$ ) shows a non-significant correlation between age and size, as does the Spearman's rank correlation coefficient ( $r_s$ )=0.51.

Finally, it should also be noted that an age–size relation within the sexes (among females: no correlation; among males: positive correlation) was not shown for the mandibular specimens (even though mandibular measurements are generally thought to show quite a high degree of sexual dimorphism in various primate and fossil hominid species).<sup>1,5,19,23</sup> Lockwood et al.<sup>18</sup> did not assign sexes to *A. robustus* mandibles, stating that 'there is greater overlap in male and female mandible size in sexually dimorphic hominoids' (p. 1445) and that 'the available specimens preserve less information than the maxillas' [Lockwood et al.<sup>18</sup> 'Supporting Online Material', p. 3]. However, in contrast to the maxillary specimens, there is a significant correlation between age and size for the whole sample of mandibular specimens ( $r_s=0.66$ ;  $p<0.05$ ,

based on data from Lockwood's Fig. 1<sup>18</sup>), which should not be the case for both sexes together, unless all of them were males.

#### Evidence for one-male polygynous groups?

##### *Sex differences in predation risk*

Lockwood et al.<sup>18</sup> argued that in sexually dimorphic primates, non-dominant adult males spend more time alone, on the periphery of a social group, or in small all-male bands.<sup>29</sup> A more peripheral, solitary life places males at greater risk of becoming victims of a predator attack.<sup>30-32</sup> An example adduced to illustrate this claim is the behaviour of *Papio cynocephalus* baboons, in which the mortality rate for dispersing males compared to males living in groups increases by at least 2–3 times.<sup>32</sup> This difference in mortality was said to match the 4-fold difference in the sex ratio, as claimed for the Swartkrans australopithecines.<sup>18</sup>

Dunbar<sup>33</sup> emphasised that although primate males frequently suffer higher mortality rates than females, this occurs at any given age, and irrespective of whether they belong to multimale or one-male groups. Besides being exposed to a higher risk of predation, males also have a greater likelihood of sustaining fatal wounds as a result of fighting over mates, or entering new groups. It was shown by Brain<sup>34,35</sup> that the bone accumulations in the Swartkrans cave can indeed, largely be attributed to predator behaviour (e.g. the Swartkrans leopard hypothesis<sup>34</sup>). Besides leopards, the primary predators suggested for the australopithecines were sabre-toothed felids and hunting hyenas.<sup>35</sup> Extant savannah baboons face threats from very predators, namely leopards, lions and hyenas,<sup>31</sup> particularly when visibility is limited. Sex differences in predation risk by large felids were investigated in these baboon populations, and while leopards were indeed shown to prey on males more frequently than on females, the actual proportion of attacked and killed males to females was about 60% vs 40%<sup>31</sup> – i.e. only 1.5 times (and not 4 times) more often.

##### *Association between delayed maturity and group composition*

Taking the gorilla as a model, Lockwood et al.<sup>18</sup> also implied a relationship between extended male growth of *A. robustus* and group composition. While an association between modes of dimorphic growth (different duration vs different rate) and different kinds of group compositions in anthropoid primates has previously been suggested,<sup>8</sup> it is the multimale, rather than the one-male species that tends to attain dimorphism through bimaturism. Although similar levels of adult body size dimorphism in gorillas and *A. robustus* might still be produced through the same ontogenetic processes, the conclusions reached regarding the same social structure of those species<sup>18</sup> are not justified. Examples of primates showing a similarly high level of bimaturism (4 years), though not quite the same group composition, are gorilla (*G. gorilla*) and rhesus monkey (*Macaca mulatta*)<sup>36</sup> – the latter composed of large multimale–multifemale social groups.

It would appear, therefore, that the Lockwood et al.<sup>18</sup> argument that difference in sex bias of male deaths resulting from predation and extended male growth be taken as evidence of one-male *Australopithecus robustus* social groups, seems dubious.

### Clues to *A. robustus* societies from ecological factors

As the arguments by Lockwood et al.<sup>18</sup> regarding the gorilla-like polygynous social structure of *Australopithecus robustus* appear to be inconclusive, one can speculate about an alternative, given the extreme sexual dimorphism of this species.<sup>18,20</sup> In doing so, it is useful to consider ecological pressures as an important factor in determining the pattern of social organisation of a species. As is well-known, even small ecological differences can bring about large consequences for social behaviour. It seems beyond doubt that the observed pronounced sexual dimorphism in the craniofacial features of *A. robustus* and the consequent dimorphism in body mass<sup>5</sup> imply either polygyny (one-male–multifemale harems) or multimale–multifemale groups with a dominance hierarchy. However, is it possible there existed uni-male 'female-defence polygyny' hominid societies on savannah?

In attempting to answer this, one additional point needs to be made. As argued by Dixon<sup>37</sup>, the social organisation of a primate group and its mating system are not necessarily synonymous. Dixon<sup>37</sup> gives the example of mountain gorilla to illustrate that major differences may exist between the two. Mountain gorilla groups may consist of up to 3–4 adult silverback males, forming therefore a multimale social group, yet the majority of copulations are performed by a single, dominant male, so that the mating system is one-male polygynous. Furthermore, in some cases, different social groups and more than one mating system does occur within the same primate species (e.g. one-male and multimale–multifemale social units in the hanuman [gray] langur, or monogamy and polyandry in some calitrichids).<sup>37</sup> This occurrence may reflect lability in the mating systems of different populations of a given species under different circumstances or environmental conditions (e.g. different food availability or density situation).<sup>38</sup>

The South African *Australopithecus robustus* occupied a different habitat to the gorillas with regard to diet, spatial distribution of resources, food availability and predation risk. Gorillas, despite the marked differences in habitats between the lowland and the mountain populations, live in forests where visibility is limited because of dense vegetation, while the ‘robust’ australopithecines lived out on the open savannah. In terms of food quality, *A. robustus* was most probably an omnivore, inclined towards herbivory, feeding on diverse plant foods requiring intensive mastication. They must have eaten seasonal fruits, but might also have significantly fed on hard food objects such as seeds, roots and nuts,<sup>39,40</sup> at least during times of food scarcity. As the southern African australopithecines appear to have enamel isotope values, indicating a considerable amount (about one-third)<sup>41</sup> of C<sub>4</sub>-derived food in their diet,<sup>42</sup> it has been suggested that savannah-based foods such as tropical grasses and sedges, and/or the animals and insects that themselves eat those plants, made up an important but variable part of its diet.<sup>41–44</sup> Others<sup>45</sup> have concluded that the C<sub>4</sub> plant underground storage organs for hominid diets was a possible source of nutrition.

There is a marked difference between the diets (staple foods) of the savannah-dwelling ‘robust’ australopithecines and the mostly tropical forest-dwelling African apes. Gorillas, despite their regional differences, primarily eat lower quality but abundant foods such as leaves, bark, bamboo shoots and herbs that they forage for mostly on the ground (lowland gorillas also forage on fruit).<sup>46–49</sup> Chimpanzees, on the other hand, are obligated fruit eaters<sup>48</sup> (for which they forage high in trees),<sup>49</sup> consuming predominantly C<sub>3</sub>-derived foods.<sup>50</sup> They are also meat-consumers.<sup>51</sup> The diet of the ‘robust’ australopithecines resembled rather more that of savannah baboons,<sup>52</sup> which consume significant quantities of C<sub>4</sub> resources such as grass seeds and roots.<sup>53</sup> If *A. robustus* competed with these terrestrial monkeys, they would have gained an advantage over the baboons in those parts of the trophic niche that contained fallback foods, which required heavy chewing and grinding.

The South African ‘robust’ australopithecines appear to have consumed food that occurred in scattered patches and their daily path lengths must therefore have been relatively large, unlike gorillas, who travel short distances (usually 1–2 km/day).<sup>9,47,49</sup> It also seems unlikely that *A. robustus* females fed dispersed from each other (as do females of the great apes) as this would have been too risky on the open savannah. For the australopithecines, being medium-sized<sup>12,13,54,55</sup> savannah inhabitants, predator pressure is not sufficiently suggestive of one-male harems because of the high risks involved. Rather, it would have been more advantageous if they were organised as cohesive multimale–multifemale groups in which extra males were of value in protecting the females and offspring. Even for much larger sized gorillas (i.e. mountain gorillas), about one-third of their social groups are multimale groups.<sup>9,56</sup> It has been suggested that the existence of multimale mountain gorilla groups increases the chances of survival and future reproductive success of all age categories of specimens, namely infants, subadults and adults.<sup>57</sup> Gorilla multimale teams can resist male outsiders more powerfully than each can achieve individually.<sup>58</sup>

In most species of primates living in groups, it is the males, not females, that tend to leave the natal groups and transfer between groups.<sup>29,30,59</sup> However, in gorillas, chimpanzees and hamadryas baboons,<sup>9,60</sup> it is the females that transfer between groups. While authors usually invoke the fundamental difference between the social structures of monkeys and apes, there are notable exceptions to this rule. For example, while savannah species of baboons: *Papio papio*, *P. anubis*, *P. cynocephalus* and *P. ursinus* form multimale–multifemale groups (based on related females), arid country *Papio hamadryas* forms one-male harems.<sup>61</sup> A few one-male units, however, frequently keep the company of one another, forming higher level social units, called clans, and several clans form bands, thus displaying a multileveled organisation of society.<sup>62,63</sup> As such, the social structure of hamadryas baboons appears to exhibit features of both types of primate social organisation: one-male and multimale. It has therefore been suggested that the multileveled organisation might be a more suitable model (than either of these two systems) for investigating the emergence of some of the aspects of early hominid societies.<sup>62</sup> Yet, on the other hand, Smuts’ observations<sup>64,65</sup> of olive baboons (*P. anubis*) are also very suggestive in the context of the evolution of male-female relationships in the human lineage. She reported that in groups of these baboons (which have multimale–multifemale societies), both females and males often develop strong preferences towards certain sexual partners in the context of long-term social bonds (which may last for years).<sup>64,65</sup>

## Concluding remarks

Ecological factors, on the one hand, and evolutionary/phylogenetic history, on the other, determine the pattern of social organisation of species<sup>66</sup> and impose limits on the range of types of this organisation that will be adaptive.<sup>38</sup> If the set of ecological variables (habitat, diet, spatial and temporal distribution of food resources, its degree of stability and predictability, and predator pressure) of hominids changed from those of their African ancestors (that which was beneficial in the forest was not so in the savannah), it can be expected that their social organisation changed too to meet the environmental changes.

I have re-examined the evidence presented by Lockwood et al.<sup>18</sup> for a polygynous, one-male harem social structure in the South African Pleistocene hominid, *Australopithecus robustus*, and have shown that the conclusions reached in the Lockwood et al. paper<sup>18</sup> can be questioned. Although *A. robustus* exhibited an increased (almost gorilla-like) degree of facial size dimorphism, I have argued that the assertions regarding evidence of a highly uneven sex ratio and the hypothesised polygynous societies of these early hominids are dubious. The approach I have tested here to determine the sex ratio in highly dimorphic species involved sorting the specimens into sexes using a *k*-means clustering technique – a method that I have demonstrated to be highly accurate for gorillas.

While Lockwood et al.<sup>18</sup> may still be correct that the mating system of *Australopithecus robustus* was one-male, I suggest that the social system of this species was more likely a multimale–multifemale one. The typical one-male harem organisation of forest-dwelling gorillas seems to be a counterproductive model for interpreting the ecology and social life of the savannah-dwelling ‘robust’ australopithecines. It is difficult to determine whether the ‘robust’ groups were based on either the kinship of females or the kinship of males. Australopithecines were unlike any of the present-day hominoids, and thus it can hardly be expected that we would be able to perfectly fit them into behavioural patterns we know today.

## Acknowledgements

I would like to thank Milford H. Wolpoff and Jan Strzalko for constructive discussions, as well as the anonymous reviewers for their useful comments on earlier drafts of the manuscript. I also thank J.F. Thackeray, formerly of the Transvaal Museum and M. Harman of the Powell-Cotton Museum, for assistance with the specimens, and D.J. Chorn for proofreading the manuscript.



## References

1. Plavcan JM. Sexual dimorphism in primate evolution. *Yrbk Phys Anthropol.* 2001;44:25–53. <http://dx.doi.org/10.1002/ajpa.10011>
2. Leutenegger W, Kelly JT. Relationship of sexual dimorphism in canine size and body size to social, behavioral and ecological correlates in anthropoid primates. *Primates.* 1977;18:117–136. <http://dx.doi.org/10.1007/BF02382954>
3. Mitani J, Gros-Louis J, Richards AF. Sexual dimorphism, the operational sex ratio, and the intensity of male competition in polygynous primates. *Am Nat.* 1995;147:966–980. <http://dx.doi.org/10.1086/285888>
4. Plavcan JM, Van Schaik CP. Interpreting hominid behavior on the basis of sexual dimorphism. *J Hum Evol.* 1997;32:345–374. <http://dx.doi.org/10.1006/jhev.1996.0096>
5. Plavcan JM. Scaling relationships between craniofacial sexual dimorphism and body mass dimorphism in primates: Implications for the fossil record. *Am J Phys Anthropol.* 2003;120:38–60. <http://dx.doi.org/10.1002/ajpa.10154>
6. Harvey PH, Kavanagh M, Clutton-Brock TH. Sexual dimorphism in primate teeth. *J Zool Lond.* 1978;186:474–485. <http://dx.doi.org/10.1111/j.1469-7998.1978.tb03934.x>
7. Martin RD, Willner LA, Dettling A. The evolution of sexual size dimorphism in primates. In: Short RV, Balaban E, editors. *The differences between the sexes.* Cambridge: Cambridge University Press; 1994. p. 159–200.
8. Leigh SR. Socioecology and the origin of sexual dimorphism in anthropoid primates. *Am J Phys Anthropol.* 1995;97:339–356. <http://dx.doi.org/10.1002/ajpa.1330970402>
9. Stewart KJ, Harcourt AH. Gorillas: Variation in female relationships. In: Smuts BB, Cheney DL, Seyfarth RM, Wrangham RW, Struhsaker TT, editors. *Primate societies.* Chicago, IL: University of Chicago Press; 1987. p. 155–164.
10. Plavcan JM. Inferring social behavior from sexual dimorphism in the fossil record. *J Hum Evol.* 2000;39:327–344. <http://dx.doi.org/10.1006/jhev.2000.0423>
11. Lockwood CA, Richmond BG, Jungers WL, Kimbel WH. Randomization procedures and sexual dimorphism in *Australopithecus afarensis*. *J Hum Evol.* 1996;31:537–548. <http://dx.doi.org/10.1006/jhev.1996.0078>
12. McHenry HM. Petite bodies of the ‘robust’ australopithecines. *Am J Phys Anthropol.* 1991;86:445–454. <http://dx.doi.org/10.1002/ajpa.1330860402>
13. McHenry HM. Body size and proportions in early hominids. *Am J Phys Anthropol.* 1992;87:407–431. <http://dx.doi.org/10.1002/ajpa.1330870404>
14. Reno PL, Meindl RS, McCollum MA, Lovejoy CO. Sexual dimorphism in *Australopithecus afarensis* was similar to that of modern humans. *Proc Natl Acad Sci USA.* 2003;100:9404–9409. <http://dx.doi.org/10.1073/pnas.1133180100>
15. Reno PL, Meindl RS, McCollum MA, Lovejoy CO. The case is unchanged and remains robust: *Australopithecus afarensis* exhibits only moderate skeletal dimorphism. A reply to Plavcan et al. (2005). *J Hum Evol.* 2005;49:279–288. <http://dx.doi.org/10.1016/j.jhev.2005.04.008>
16. Plavcan JM, Lockwood CA, Kimbel WH, Lague MR, Harmon EH. Sexual dimorphism in *Australopithecus afarensis* revisited: How strong is the case for a human-like pattern of dimorphism? *J Hum Evol.* 2005;48:313–320. <http://dx.doi.org/10.1016/j.jhev.2004.09.006>
17. Lee S-H. Patterns of size sexual dimorphism in *Australopithecus afarensis*: Another look. *Homo.* 2005;56:219–32. <http://dx.doi.org/10.1016/j.jchb.2005.07.001>
18. Lockwood CA, Menter CG, Moggi-Cecchi J, Keyser AW. Extended male growth in a fossil hominin species. *Science.* 2007;318:1443–1446. <http://dx.doi.org/10.1126/science.1149211>
19. Lockwood CA. Sexual dimorphism in the face of *Australopithecus africanus*. *Am J Phys Anthropol.* 1999;108:97–127. [http://dx.doi.org/10.1002/\(SICI\)1096-8644\(199901\)108:1<97::AID-AJPA6>3.0.CO;2-0](http://dx.doi.org/10.1002/(SICI)1096-8644(199901)108:1<97::AID-AJPA6>3.0.CO;2-0)
20. Kaszycka KA. Dymorfizm płciowy południowoafrykańskich australopiteków [Sexual dimorphism in the South African australopithecines]. Poznan: Wyd. Naukowe UAM; 2009. Polish.
21. Gibbons A. Hominid harems: Big males competed for small australopithecine females. *Science.* 2007;318:1363. <http://dx.doi.org/10.1126/science.318.5855.1363a>
22. Wolpoff MH. Sexual dimorphism in the australopithecines. In: Tuttle RH, editor. *Paleoanthropology: morphology and paleoecology.* The Hague: De Gruyter Mouton; 1975. p. 245–284. <http://dx.doi.org/10.1515/9783110810691.245>
23. Wolpoff MH. Some aspects of the evolution of early hominid sexual dimorphism. *Curr Anthropol.* 1976;17:579–606. <http://dx.doi.org/10.1086/201798>
24. Hartigan JA, Wong MA. Algorithm AS 136 A K-means clustering algorithm. *Appl Stat J Roy St C.* 1979;28:100–108. <http://dx.doi.org/10.2307/2346830>
25. Kaszycka KA. Status of Kromdraai: Cranial, mandibular and dental morphology, systematic relationships, and significance of the Kromdraai hominids. Paris: CNRS; 2002. p.130
26. Plavcan JM. Comparison of four simple methods for estimating sexual dimorphism in fossils. *Am J Phys Anthropol.* 1994;94:465–476. <http://dx.doi.org/10.1002/ajpa.1330940403>
27. Keyser AW. The Drimolen skull: The most complete australopithecine cranium and mandible to date. *S Afr J Sci.* 2000;96:189–193.
28. Morissette L, Chartier S. The k-means clustering technique: General considerations and implementation in Mathematica. *Tutorials Quant Meth Psych.* 2013;9:15–24.
29. Pusey AE, Packer C. Dispersal and philopatry. In: Smuts BB, Cheney DL, Seyfarth RM, Wrangham RW, Struhsaker TT, editors. *Primate societies.* Chicago, IL: University of Chicago Press; 1987. p. 250–266.
30. Harcourt AH. Strategies of emigration and transfer by primates, with particular reference to gorillas. *Z Tierpsychol.* 1978;48:401–420. <http://dx.doi.org/10.1111/j.1439-0310.1978.tb00267.x>
31. Cowlshaw G. Vulnerability to predation in baboon populations. *Behaviour.* 1994;131:293–304. <http://dx.doi.org/10.1163/156853994X00488>
32. Alberts SC, Altmann J. Balancing costs and opportunities: Dispersal in male baboons. *Am Nat.* 1995;145:279–306. <http://dx.doi.org/10.1086/285740>
33. Dunbar RIM. Demography and reproduction. In: Smuts BB et al., editors. *Primate societies.* Chicago, IL: University of Chicago Press; 1987. p. 240–249.
34. Brain CK. New finds at the Swartkrans australopithecine site. *Nature.* 1970;225:1112–1119. <http://dx.doi.org/10.1038/2251112a0>
35. Brain CK. The hunters or the hunted. An introduction to African cave taphonomy. Chicago, IL: University of Chicago Press; 1981.
36. Leigh SR. Patterns of variation in the ontogeny of primate body size dimorphism. *J Hum Evol.* 1992;23:27–50. [http://dx.doi.org/10.1016/0047-2484\(92\)90042-8](http://dx.doi.org/10.1016/0047-2484(92)90042-8)
37. Dixon AF. Primate sexuality. Comparative studies of the prosimians, monkeys, apes, and human beings. Oxford: Oxford University Press; 1998.
38. Emlen ST, Oring LW. Ecology, sexual selection, and the evolution of mating systems. *Science.* 1977;197(4300):215–223. <http://dx.doi.org/10.1126/science.327542>
39. Grine FE. Dental evidence for dietary differences in *Australopithecus* and *Paranthropus*: A quantitative analysis of permanent molar microwear. *J Hum Evol.* 1986;15:783–822. [http://dx.doi.org/10.1016/S0047-2484\(86\)80010-0](http://dx.doi.org/10.1016/S0047-2484(86)80010-0)
40. Kay RF, Grine FE. Tooth morphology, wear and diet in *Australopithecus* and *Paranthropus* from southern Africa. In: Grine FE, editor. *Evolutionary history of the ‘robust’ australopithecines.* New York: Aldine de Gruyter; 1988. p. 427–447.
41. Lee-Thorp JA, Sponheimer M. Contributions of biogeochemistry to understanding hominin dietary ecology. *Year Phys Anthropol.* 2006;49:131–148. <http://dx.doi.org/10.1002/ajpa.20519>
42. Lee-Thorp JA, Van der Merwe NJ, Brain CK. Diet of *Australopithecus robustus* at Swartkrans from stable carbon isotopic analysis. *J Hum Evol.* 1994;27:361–372. <http://dx.doi.org/10.1006/jhev.1994.1050>
43. Sponheimer M, Alemseged Z, Cerling TE, Grine FE, Kimbel WH, Leakey MG, et al. Isotopic evidence of early hominin diets. *P Natl Acad Sci USA.* 2013;110(26):10513–10518. <http://dx.doi.org/10.1073/pnas.1222579110>

44. Sponheimer M, Passey BH, De Ruiter DJ, Guatelli-Steinberg D, Cerling TE, Lee-Thorp JA. Isotopic evidence for dietary variability in the early hominin *Paranthropus robustus*. *Science*. 2008;314:980–982. <http://dx.doi.org/10.1126/science.1133827>
45. Yeakel JD, Bennett NC, Koch PL, Dominy NJ. The isotopic ecology of African mole rats informs hypotheses on the evolution of human diet. *Proc R Soc B*. 2007;274(1619):1723–1730. <http://dx.doi.org/10.1098/rspb.2007.0330>
46. Rogers ME, Abernethy K, Bermejo M, Cipolletta C, Doran D, McFarland K, et al. Western gorilla diet: A synthesis from six sites. *Am J Primatol*. 2004;64:173–192. <http://dx.doi.org/10.1002/ajp.20071>
47. Doran-Sheehy DM, Greer D, Mongo P, Schwindt D. Impact of ecological and social factors on ranging in western gorillas. *Am J Primatol*. 2004;64:207–222. <http://dx.doi.org/10.1002/ajp.20075>
48. Yamagiwa J, Basabose AK. Diet and seasonal changes in sympatric gorillas and chimpanzees in Kahuzi-Biega National Park. *Primates*. 2006;47:74–90. <http://dx.doi.org/10.1007/s10329-005-0147-7>
49. Stanford CB. The behavioral ecology of sympatric African apes: Implications for understanding fossil hominoid ecology. *Primates*. 2006;47:91–101. <http://dx.doi.org/10.1007/s10329-005-0148-6>
50. Sponheimer M, Loudon JE, Codron D, Howells ME, Pruett JD, Codron J, et al. Do 'savanna' chimpanzees consume C4 resources? *J Hum Evol*. 2006;51:128–133. <http://dx.doi.org/10.1016/j.jhevol.2006.02.002>
51. Stanford CB, Wallis J, Matama H, Goodall J. Patterns of predation by chimpanzees on red colobus monkeys in Gombe National Park, Tanzania, 1982–1991. *Am J Phys Anthropol*. 1994;94:213–228. <http://dx.doi.org/10.1002/ajpa.1330940206>
52. Strum SC. Processes and products of change: Baboon predatory behavior at Gilgil, Kenya. In: Harding R, Teleki G, editors. *Omnivorous primates: Gathering and hunting in human evolution*. New York: Columbia University Press; 1981. p. 255–302.
53. Codron D, Lee-Thorp JA, Sponheimer M, De Ruiter D, Codron J. Inter- and intrahabitat dietary variability of chacma baboons (*Papio ursinus*) in South African savannas based on fecal  $d^{13}C$ ,  $d^{15}N$ , and %N. *Am J Phys Anthropol*. 2006;129:204–214. <http://dx.doi.org/10.1002/ajpa.20253>
54. Steudel K. New estimates of early hominid body size. *Am J Phys Anthropol*. 1980;52:63–70. <http://dx.doi.org/10.1002/ajpa.1330520109>
55. Spocter MA, Manger PR. The use of cranial variables for the estimation of body mass in fossil hominins. *Am J Phys Anthropol*. 2007;134:92–105. <http://dx.doi.org/10.1002/ajpa.20641>
56. Harcourt AH, Stewart KJ. Gorilla society: What we know and don't know. *Evol Anthropol*. 2007;16:147–158. <http://dx.doi.org/10.1002/evan.20142>
57. Robbins MM. A demographic analysis of male life history and social structure of mountain gorillas. *Behaviour*. 1995;132:21–47. <http://dx.doi.org/10.1163/156853995X00261>
58. Strier KB. Beyond the apes: Reasons to consider the entire primate order. In: De Waal FBM, editor. *Tree of origin. What primate behaviour can tell us about human social evolution*. Cambridge, MA: Harvard University Press; 2002. p. 70–93.
59. Packer C. Inter-troop transfer and inbreeding avoidance in *Papio anubis*. *Anim Behav*. 1979;27:1–36. [http://dx.doi.org/10.1016/0003-3472\(79\)90126-X](http://dx.doi.org/10.1016/0003-3472(79)90126-X)
60. Wrangham RW. An ecological model of female-bonded primate groups. *Behaviour*. 1980;75:262–300. <http://dx.doi.org/10.1163/156853980X00447>
61. Stambach E. Desert, forest and montane baboons: Multilevel-societies. In: Smuts BB, et al., editors. *Primate societies*. Chicago, IL: University of Chicago Press; 1987. p. 112–120.
62. Sigg H, Stolba A, Abegglen J-J, Dasser V. Life history of hamadryas baboons: Physical development, infant mortality, reproductive parameters and family relationships. *Primates*. 1982;23:473–487. <http://dx.doi.org/10.1007/BF02373959>
63. Abegglen J-J. On socialization in hamadryas baboons. A field study. Cranbury: Associated University Press; 1984.
64. Smuts BB. Sex and friendship in baboons. New York: Aldine; 1985.
65. Smuts BB. Sexual competition and mate choice. In: Smuts BB, Cheney DL, Seyfarth RM, Wrangham RW, Struhsaker TT, editors. *Primate societies*. Chicago, IL: University of Chicago Press; 1987. p. 385–399.
66. Wrangham RW. Evolution of social structure. In: Smuts BB, Cheney DL, Seyfarth RM, Wrangham RW, Struhsaker TT, editors. *Primate societies*. Chicago, IL: University of Chicago Press; 1987. p. 282–296.



# Visitors' views of human origins after visiting the Cradle of Humankind World Heritage Site

**AUTHOR:**

Anthony Lelliott<sup>1</sup>

**AFFILIATION:**

<sup>1</sup>Wits School of Education,  
University of the Witwatersrand,  
Johannesburg, South Africa

**CORRESPONDENCE TO:**

Anthony Lelliott

**EMAIL:**

tony.elliott@wits.ac.za

**POSTAL ADDRESS:**

Wits School of Education,  
University of the Witwatersrand,  
Private Bag 3, Wits 2050,  
South Africa

**DATES:**

**Received:** 31 May 2015

**Revised:** 09 July 2015

**Accepted:** 04 Aug. 2015

**KEYWORDS:**

scientific evidence;  
human origins; ancestor;  
evolution; palaeontology

**HOW TO CITE:**

Lelliott A. Visitors' views of human origins after visiting the Cradle of Humankind World Heritage Site. *S Afr J Sci.* 2016;112(1/2), Art. #2015-0210, 8 pages.  
<http://dx.doi.org/10.17159/sajs.2016/20150210>

The Cradle of Humankind World Heritage Site, west of Johannesburg, was designated in 1999 because of its importance as a locality where numerous hominid fossils have been discovered since the 1930s. In this article, responses to questions from a survey of more than 800 adult visitors to the Cradle of Humankind visitor centres are analysed, covering their understanding of the concept of the 'cradle' and their views on human evolution. Findings indicated that 63% of the respondents conceptualised the cradle as the origin or birthplace of humankind, and a similar proportion thought that nowhere else could be called the Cradle of Humankind (77% of people of South African nationality thought this). Nearly 60% of respondents accepted that humans evolved from an ape-like ancestor, while 25% disagreed. South Africans were less likely to accept human evolution than their international counterparts. The great majority of participants who accepted human evolution based their agreement on various forms of evidence and their knowledge of evolution. A religious foundation was used for their rationale by 60% of those who rejected evolution, with 33% citing evidence for their rejection. The implications of the findings are discussed in the light of public awareness and human origins.

## Introduction

It is widely accepted amongst biologists and biology educators that evolution is a crucial organising principle on which much of the field of biology is based.<sup>1</sup> Despite this conviction, there is considerable evidence that public non-acceptance of evolution holds sway in many countries because of a combination of religious beliefs and difficulties in understanding the principles on which the theory rests. Studies demonstrate that considerable proportions of the population hold cultural and religious views which they consider exclude their acceptance of the theory of evolution.<sup>2</sup> Similarly, studies show that the concepts of variation, inheritance, natural selection and descent with modification are difficult to understand, even when explicitly taught in school.<sup>3,4</sup> In this paper, an analysis of the views of over 800 visitors to the Cradle of Humankind is provided. The Cradle of Humankind is a World Heritage Site designated in 1999, which is home to a large proportion of the world's fossil hominids (the revised classification 'hominins' is not used here).

The research findings described in this paper are significant for South Africa for a number of reasons. Firstly, 'African Origins' is a focus area of the Department of Science and Technology which covers interdisciplinary research programmes in palaeontology, archaeology and genetics, and includes the evolution of humankind. The findings will inform such programmes, so that they can be more carefully tailored to the existing knowledge of the public, rather than relying on surveys from overseas. The notion of a 'Cradle of Humankind' relates both to Africa as the birthplace of early human ancestors (early hominids) as well as the birthplace of modern humans. How such a concept of 'birthplace' has crossed from the scientific community to the general public is worthy of examination – is it taken literally or is it used as a metaphor? Secondly, new fossil finds continue to be made in the Cradle. Understanding what visitors think about human origins is important in this process and the insights described in this paper provide valuable background material for public engagement. There are no studies in South Africa on the views of adults concerning human origins, a gap which the paper fills. Thirdly, the survey undertaken in the paper identifies some of the numerous misconceptions which visitors possess with respect to human evolution. Addressing such misconceptions should be one of the roles of the visitor centres in the Cradle of Humankind.

The research questions that the study examined are:

- How do visitors to the Cradle of Humankind understand the term 'cradle'?
- To what extent do visitors to the Cradle of Humankind accept the concept of human evolution, and how do they rationalise their view?

A conceptual model inspired by researchers in the field of evolution in the museum context was used during data analysis.<sup>5,6</sup> This model is based on reasoning patterns of visitors giving explanations of their verbal responses to interview questions about the evolution of organisms, and was originally devised as a tool for both teachers and researchers in the field of evolution.<sup>5</sup> The model was adapted for the purposes of this study and replaces 'reasoning' with 'rationalising statements' for two reasons. First, unlike the museums which were part of the Evans et al.<sup>5</sup> research, the Cradle of Humankind visitor centres do not explicitly 'teach' an understanding of the process of evolution. Secondly, the data collected were in the form of written responses, rather than reasoning during verbal interviews.

## Context of the study

Prior to 1994, the South African political system was based on apartheid ideology. One of the central tenets of apartheid was its adherence to Christian National Education,<sup>7</sup> and by the 1960s, there was no reference to Darwin or evolution in the school curricula. The result was that at least two generations of South Africans were denied access to this important aspect of scientific literacy.<sup>8</sup> Ironically, discoveries of early hominid fossils were being made in the

area that would later become the Cradle of Humankind throughout the 20th century. However, the great majority of the South African population would have had no way of appreciating such discoveries in the unfolding human story, through formal schooling.

The original idea that Africa was the 'birthplace' of humans originated with Charles Darwin, who, based on chimpanzees and gorillas occurring in Africa stated, 'It is somewhat more probable that our early progenitors lived on the African continent than elsewhere'.<sup>9</sup> Although scientists favoured Europe or Asia in the early 20th century, as more fossil evidence was uncovered in south and east Africa, the concept of humans originating in Africa became accepted by most Western palaeoanthropologists.<sup>10</sup> However, the notion of modern humans developing in Africa and migrating to all other parts of the globe<sup>11</sup> is still contested in some quarters.<sup>12</sup> The findings below show how people's notions of 'Cradle' and African origins have crossed into the public domain.

The context of the study is clearly within the field of 'informal environments' and specifically 'designed settings'<sup>13</sup> in which the visitor centre space has been devised to provide information about the palaeontological findings in the area and their significance. While scientific findings have been published from the Cradle of Humankind and elsewhere, literature research for this article has revealed there is a paucity of published research on museum education relating to science in the country. Overseas, the situation is very different. In the informal sector in the USA, Storksdieck and Stein<sup>14</sup> found that approximately half of museum audiences accepted the scientific explanation for human origins compared to only 27% of the national sample. However, creationist views are also often endorsed for human origins, and many museum visitors have difficulty in understanding the processes of evolution and hold alternative conceptions about it.<sup>15</sup> Some small-scale studies have shown that museums can contribute to changing visitors' views about evolution. In their first study, Spiegel and colleagues showed that, depending on the type of organism they were considering, visitors reasoned differently about evolution. Visitors tended to combine evolutionary knowledge with intuitive reasoning or, less often, creationist reasoning.<sup>6</sup> More recently, Spiegel found that a single visit to an interactive exhibition on evolution resulted in small changes in visitors' scientific understanding of evolution, regardless of their existing views towards it.<sup>16</sup>

Scott<sup>17-19</sup> has written extensively on the various ways in which visitors make meaning while visiting exhibitions which explain human origins. She found that visitors' 'entrance narratives'<sup>20</sup> had a strong influence on visitors' perceptions of what the museums were portraying. Current ideas about hominid ancestors are becoming increasingly complex,<sup>21</sup> and how such concepts are understood by museum visitors is of growing importance. Scott<sup>18</sup> suggests that museums are being increasingly challenged in the ways they portray human evolution: they need to provide significant experiences for visitors while at the same time devising novel approaches to explain the theories, methodologies and evidence from new hominid evidence from the field. Such learning in informal settings might include aspects of the characteristics of science, such as the tentative nature of scientific knowledge, the status of a theory, and suchlike. Given the very low level of scientific literacy in South Africa<sup>22</sup>, designed settings such as the Cradle of Humankind visitor centres need to educate as well as entertain.

Within the Cradle of Humankind, there are two visitor centres open to the public. Maropeng ('returning to the place of origin' in the SeTswana language) provides wide-ranging exhibits for the mostly self-guided visitor. These include earth history, palaeontology, how pre-human ancestors have developed into modern humans, as well as information about human impact on the planet. The centre includes features associated with a theme park, such as a boat ride twisting through an artificial cave and a walk across a rotating tunnel (the 'vortex') representing the formation of the universe. In contrast, the Sterkfontein Visitor Centre provides a smaller museum section showing human ancestral development, and a guided tour through underground caves from which hominid fossils have been excavated. Sterkfontein does not include the features of a theme park found at Maropeng. The official website of the Cradle of Humankind ([maropeng.co.za](http://maropeng.co.za)) cites the main

purposes of the facilities as being for tourism development, protection and management of the excavated heritage sites and for scientific research. Although there is no explicit link to education, both visitor centres play host to thousands of schoolchildren annually during school excursions.

## Method

The approach adopted for this study was a post-visit only design because it was not an impact evaluation. Instead, it aimed to capture participants' views of two aspects of human origins within the context of a visit to the Cradle of Humankind. Data collection consisted of a survey using continual ask sampling,<sup>23</sup> where visitors were intercepted by a researcher as they exited the centre and invited to participate in the study. This type of sampling is considered best practice in informal learning contexts, and can be regarded as a close approximation to a random sample.<sup>24</sup> Once they had agreed to participate, the individuals or pairs completed the questionnaire themselves. The surveys were conducted over 19 days between April and July 2013. Data logs of refusal rates were kept, and using total ticket sales for each day, it was calculated that the data collected ( $n=437$  at Maropeng;  $n=374$  at Sterkfontein) represented a mean 16% ( $SD=7.1$ ) of daily visitors. The sample was not representative of the South African population, but was a self-selected group of individuals and groups who decided to visit the Cradle of Humankind.

A questionnaire was developed using questions designed to elicit responses from visitors, based on ideas developed by Scott<sup>19</sup>, as well as other evaluative tools such as those of Falk and Storksdieck<sup>25</sup>. The first section of the questionnaire asked demographic questions: age, gender, occupation, nationality and the type of group the visitors were in. An optional question on 'ethnic background' was included. The term 'ethnic background' was chosen as a signifier for 'population group' which is used by Statistics South Africa in its census surveys. Affirmative action policies are current in South Africa society, and statistics on race/ethnicity are kept for redress purposes. A questionnaire was used for pragmatic reasons, in order to survey the largest number of visitors in a relatively short time period without inconveniencing them too much. An interview would have taken more of their time and the refusal rate would likely have been higher, although richer data from fewer participants would probably have resulted. It is accepted that questionnaires have their weaknesses. These include non- or partial-completion, untruthful completion, bias or lack of thoughtfulness in who completes them and so forth.<sup>26</sup> Both Scott<sup>19</sup> and Dickinson<sup>27</sup> noted the limitations of collecting survey data in informal settings, but in the current study, the advantage of collecting a snapshot of over 800 visitors' views was considered fit for purpose.

The questionnaire was face- and content-validated by two experts in evolution education, and piloted prior to data collection commencing. Adjustments were made to the questions, so that they were regarded as valid for the purpose of the study. Ethical clearance for the study was obtained from the Human Research Ethics Committee (Non-Medical) of University of the Witwatersrand (Protocol number 2013ECE019S); all participants were provided with information sheets and they signed consent forms.

In this article, participant responses to three questions from the questionnaire were examined. Questions 8 and 9 asked about the term Cradle of Humankind: why the area had been given this name, and what other area could be so called. This question was developed for two reasons. The first reason was in relation to the question that Scott<sup>19</sup> asked of museum visitors: 'Do you think of Africa as the Cradle of Mankind?' The second reason was to determine participants' contextual understanding of the term 'cradle', the rationale for its use, as well as their knowledge about early human ancestors elsewhere (e.g. east Africa).

Question 10 was, 'Do you accept that humans evolved from an ape-like ancestor? Y/N; explain your thinking in giving this answer'. This was devised to ascertain the visitors' views on human evolution, and was worded to try to avoid the notion of 'belief'. However, 'evolved from' and 'ape-like ancestor' are loaded terms and tend to elicit a response



informed by the visitors' entrance narrative rather than anything they experienced during their visit.<sup>19</sup> General usage of the term 'ancestor' refers to a predecessor of current living persons, and is likely to be interpreted as such by participants. Biologically, 'ancestor' refers to a progenitor, from which an organism has developed or descended. The formulation of question 10 was difficult; Media24<sup>28</sup> used the phrase 'developed from previous species of animals' (the original question was in Afrikaans). The Pew Research Center<sup>29</sup> used the phrase 'humans and other living things have evolved over time', but this does not get to the heart of the issue regarding what humans evolved *from*. So, despite its flaws, question 10 was asked in the form shown.

A quantitative analysis was conducted using frequency counts (descriptive statistics). Where multiple answers were given by the same respondent, the first answer given was used in the analysis, except where indicated below. All inferential statistical findings discussed below used the chi-squared test. Qualitative analysis of the open-ended responses to Questions 8 and 10 were coded inductively.<sup>30</sup> In question 8, a hierarchy of codes was established, whereby, for example, human (humankind, mankind, man etc) > hominid > fossil > skeleton, so that if a respondent referred to human and fossil and skeleton in their answer, it was coded *human*. Inter-coder reliability is important in coding open-ended questions from questionnaires.<sup>31</sup> Reliability was established by getting a second coder to recode the data using a codebook. A randomly selected 10% of the second coder's data indicated an inter-coder reliability of 75% (Cohen's kappa = 0.75), which indicates good agreement.

Rationalising statements derived from the Spiegel and colleagues' conceptual model<sup>16</sup> used during analysis of the qualitative data, are shown in Table 1. To exemplify how the model was used, the right hand column shows an example of each rationalising statement in response to Question 10 in the questionnaire. The statements are used as categories of responses, while the data covered in the findings reflect the responses made by the participants themselves.

## Findings

### Sample

Broad demographic characteristics of the visitors were as follows: 49% of the visitors were in their 20s or 30s, and 27% in the 40–60 age range. Of the sample, 45% were women, 40% men, 8% completed the questionnaire as a couple, while 7% did not state their gender. Of the visitors, 64% were from South Africa, 12% from the Americas, 11% from Europe, 5% from Asia, and 3% each from Oceania and Africa. Most of the visitors (40%) were part of a family group, while 25% were couples, 12% were individuals, 11% were part of a tour group and the remainder were friends or colleagues. Of the sample, 73% followed a professional

career, 8% were students, while the remaining 19% did not state their occupation. The optional question on ethnic origin was answered by 72% of participants, and of those who did, 62% were white, 24% were black and the remainder were of Indian (9%) or Asian (2%) descent or mixed race (2%).

Question 8 read as follows: 'The area of Gauteng/North West Province in which Maropeng lies is sometimes called the 'Cradle of Humankind'. Why do you think it is called this?'

Of the 659 responses to Question 8, 63% referred to the 'origin' or 'birthplace' of 'humans', 'humankind' or 'mankind' (see Table 2). The following quotations are a selection of responses to the question (index numbers in brackets refer to codes of individual quotations, together with some demographic information):

- 'Humankind started here' (1/5 22: South African female; 50s)
- 'The birthplace of mankind – our common birthplace' (1/5 32: French female; 50s)
- 'Because this is where most human remains have been found of the early human beings' (1/5 13: South African female; 20s)

One would expect these sorts of answers, as the meaning of the term can be deduced from the words 'Cradle of Humankind'. Only a few participants (about 1%) expressed scepticism about the name 'Cradle'. For example:

- 'Wishful thinking/marketing – the actual origins of humankind is surely not so precisely known' (11/5 16: British male; 40s)
- 'Fanciful claim based on the finds in several local caves' (26/5 38: Australian male; 70s).

While the majority of participants referred to the concept of the cradle being the origin of humankind, 11% pointed to the presence of hominid fossils in the area, and a further 10% used the term fossil, but did not specify what sort of fossils had been found. Small proportions used the term 'skeleton' (3%), *H. sapiens* (2%) and 'ancestor' (1%) without also including the term 'human', 'hominid' or similar.

There was no statistically significant relationship ( $p=0.379$ ) between the answers to Question 8 and whether or not participants accepted human evolution (Question 10). However, those respondents who accepted evolution were significantly statistically more likely to refer to 'hominids' (13%,  $p<0.05$ ) in their answer to Question 8 than respondents who did not accept evolution (6%,  $p<0.05$ ). One might speculate that the latter were less familiar with the term hominid, or they were less

Table 1: Conceptual model of rationalising statements

Rationalising statement	Description	Example of response
Naturalistic evidential	Any reference to evidence or similarities	'The fossil record combined with DNA evidence combined with anthropological research is hard to argue with.'
Naturalistic evolution	Reference to evolution only, but not to evidence	'I just believe in the theory of evolution and the fittest will survive.'
Naturalistic scientific	An appeal to 'science' or 'logic', without further explanation	'There is no better proved theory that explains the origins of humankind. Such a theory is logical and realistic.'
Creationist religious	Rationale refers to religious belief (God, Allah, Bible, etc.)	'God created man and animals, one did not evolve to the other.'
Creationist evidential	Rationale discusses evidence but does not refer to religion overtly	'Because humans are not evolving still. I see no evidence of this. We remain human!!'
Mixed rationale	Uses a combination of naturalistic and creationist statements	'Science as seen here indicates that there was evolution. The bible says God creates man.'

accepting of hominids as human ancestors when discussing the Cradle of Humankind.

**Table 2:** Frequency of codes for participants explaining the name 'Cradle of Humankind' ( $n=659$ )

Codes	Frequency %
Human	63
Hominid	11
Fossil	10
Other	4
Skeleton	3
Origin of us	3
Origin of life	2
Sapiens	2
Ancestor	1
Sceptical view	1

Question 9 reads: 'To your knowledge, is there anywhere else that is or could be called the Cradle of Humankind? (Yes/No). If you answered "yes" to Question 9, Where is it? Why is it called the Cradle of Humankind?' In answering the first part of this question ( $n=811$ ), 63% of the participants thought that there was *nowhere else* that is or could be called the Cradle of Humankind. Only 22% of respondents considered that there might be another place which could be so named (while 15% did not answer the question). These figures suggest that the respondents were unaware of claims to the name which could be made by both Kenya and Ethiopia. Of those who answered the question ( $n=612$ ), there was a statistically significant relationship between respondents being aware of 'another Cradle' and nationality (chi-squared  $p=0.000598$ ; effect size Cramér's  $V=0.197$ , i.e. low to medium). South Africans ( $n=397$ ) were more likely to *disagree* with the statement ('no' 77%), while Americans, the next highest regional grouping who completed the survey ( $n=77$ ), showed a significantly lower frequency of disagreement ('no' 60%). Although non-South African Africans were few in number, the majority (10/15) agreed with the statement.

Of the small number of participants ( $n=135$ ) who suggested where else a Cradle of Humankind might be, 56% referred to east Africa or one of its constituents (Kenya, Ethiopia or Tanzania), while a further 15% listed 'Africa' or another African country (e.g. Chad, Nigeria). Just over 10% mentioned another part of South Africa, while a few respondents cited the Middle East, Far East or Europe. While some of these areas might be guesses, a minority of the visitors to the Cradle of Humankind were aware of other areas of Africa and the world which are prominent in the study of human origins.

There were three other items of interest in the responses to Questions 8 and 9. A total of 15 respondents (2%) appeared to conflate the notion

of the 'origin of humankind' with the 'origin of life', as they used terms such as the following in answer to Question 8:

- 'This is the place where we have found evidence of origin of life' (18/5 21: Indian female; teenager)
- 'It's the place where life originated' (2/6 25: South African female; 20s)

Others talked about the cradle as the place where 'it' all started:

- 'Where it all began' (26/5 40: South African female; 30s & 2/6 27: South African male; 20s)
- 'It's where everything started' (14/7 3: South African female; 30s)

Although the questionnaire instrument did not allow for elaboration of these responses, the responses suggest that these participants do not distinguish between the origin of humans and the origins of life. This misconception has implications for how such issues are presented at museums and visitor centres, and is discussed below.

A total of 5 respondents referred to the Out of Africa (OOA) theory, which states that anatomically modern humans evolved in Africa and spread from there throughout the world. One of the exhibits at Maropeng Visitor Centre explains this, and contrasts it with the Multiregional Theory for human development. However, 'Out of Africa' is a vague notion, and may refer either to OOA I, in which early hominids spread across the world, as cited by this respondent: 'This is where the first hominid stood up and started its journey toward Asia and Europe' (1/5 31). Alternatively, OOA II refers to the spread of modern humans, thus: 'It is where we, modern humans are believed to have developed and migrated from' (1/5 83: British male; 60s).

Considering that 25% of the participants did not accept that humans evolved from an ape-like ancestor (see below), only four people discussed 'creation' in their response to Question 8 or showed scepticism towards evolution. The paucity of discussion of creation and scepticism towards evolution suggests that the great majority of respondents do not dispute the cradle narrative of human origins, hominids and fossils. This may result from the structure of the questionnaire, which asks relatively uncontroversial questions to this point. Once the term evolution was introduced in Question 10, non-acceptance of scientific explanations of human origins became more outspoken.

Question 10 in the survey was, 'Do you accept that humans evolved from an ape-like ancestor?' Y/N; Explain your thinking in giving this answer. Of the respondents, 59% ( $n=811$ ) accepted the concept, 25% disagreed, while 16% did not state their opinion or gave another answer.

There were no statistically significant relationships between acceptance of human evolution and gender, age or occupation (confidence levels were set at 95%). Of the respondents who stated their nationality and answered the question ( $n=705$ ), there was a statistically significant relationship between acceptance of evolution and nationality ( $p<0.00001$ ; effect size medium: Cramér's  $V=0.221$ ). Although more South Africans agreed than disagreed with the statement, this figure (59%) was significantly lower than expected ( $p<0.001$ ), and their disagreement (37%) was higher than expected ( $p<0.001$ ). Significantly more people from the Americas and Europe agreed (89%,  $p<0.001$ ), with fewer than 10% disagreeing ( $p<0.001$ ). The data are shown in Table 3, and suggest that the South Africans who participated in the

**Table 3:** Percentage acceptance of human evolution by nationality ( $n=705$ )

	South Africa	Americas	Europe	Asia	Oceania	Africa (non-South African)	Multiple
Yes	58.6	89.2	89.5	70.3	100.0	55.0	86.7
No	37.4	7.2	9.2	29.7	0.0	40.0	13.3
Other	4.0	3.6	1.3	0.0	0.0	5.0	0.0

survey are less likely than their international counterparts to accept human evolution. Possible reasons for this will be investigated in a later phase of the study involving interviews.

As shown in Table 4, of those who stated their ethnic background ( $n=578$ ), there was a statistically significant relationship between acceptance of evolution and population group ( $p=0.0015$ ), though the effect size was small (Cramér's  $V=0.145$ ). Compared to the other groups, black participants were significantly less likely to answer 'yes' to the question (44%  $p<0.001$ ), and significantly more likely to *not* answer it (20%  $p<0.001$ ). These data suggest that black South Africans are less accepting of human evolution than other population groups. This may be related to religious beliefs<sup>32</sup> not explored here, and will be investigated further by interviewing participants. Although the data from both Tables 3 and 4 were affected by sampling, interviews will also be able to determine the interest shown in the Cradle of Humankind by people who do not accept evolution.

**Table 4:** Percentage acceptance of human evolution by population group ( $n=578$ )

	White	Black	Indian	Asian	Mixed race	Hispanic
Yes	63.6	43.8	69.2	66.7	83.3	80
No	24.7	29.9	17.3	25	16.7	0
Not answered	8.9	20.4	13.5	8.3	0	0
Other	2.8	5.8	0	0	0	20

In their study of 32 visitors to the Explore Evolution exhibition on display in Midwestern museums in the USA, Spiegel and colleagues<sup>6</sup> found that their respondents used mixed patterns of reasoning to explain evolutionary events. In the current study, the most interesting answers in the questionnaire came from the participants' responses to Question 10: 'Explain your thinking in giving [your] answer'. A total of 68% ( $n=551$ ) of the respondents gave an explanation, and the following findings refer to that proportion of the participants who responded, recalculated to 100% (see Table 5). The respondents' explanations fell into the two clear groups of agreement or disagreement, with some expanding on their 'maybe' or 'yes and no' answers. Both groups contained examples of misconceptions in their statements.

Of those who agreed with the statement ( $n=362$ ) in Question 10, the largest combined category (47%,  $n=171$ ) was the respondents who cited various forms of evidence to back up their assertion. This included anatomical, genetic, fossil, and behavioural facts in support of evolution, for example:

- 'The ape-like ancestors had many physical features that evolved into current human species – based on mode of erect mobility and use of the "thumb"' (2/6 30: South African female; 60s)

- 'Genetic makeup is similar to man' (11/5 28: South African female; 30s)
- 'Fossils give an indication as to where we evolved from' (16/6 26: South African female; 20s)
- 'We resemble many behaviours of theirs' (9/6 23: Indian male; 20s)

Some of these explanations cited 'similarities' between humans and apes (or ape-like ancestors), but did not specify what such similarities are, for example:

- 'similarity in looks' (18/5 10: Indian female; 20s)
- 'It has always been in my mind as to why the resemblance was so much!' (13/7 38: South African male; 20s)

The next largest category (27%,  $n=96$ ) was those respondents who referred to evolution in their explanation. Comments ranged from simple one-word answers or phrases: 'Evolution', 'Origin of the species' to more detailed explanations covering beliefs:

- 'I believe in the theory of evolution' (1/5 71: South African male; 40s)

And elements of reasoning:

- 'If you look at apes today you can still see in some ways how they have evolved to become humans – just how time progressed and evolution set in' (misconception; 2/6 32: South African female; 20s)

Other explanations referred to evidence, but did not actually specify what the evidence was, for example:

- 'Evidence' (1/5 1 & 9/6 8: both South African male; 30s)
- 'The evidence overwhelmingly points in that direction' (1/5 86: Irish male; 30s)

Of the explanations, 18% ( $n=64$ ) appealed to 'science' or 'logic', with little further explanation. For example:

- 'Scientific evidence' (11/5 26: Tanzanian male; 30s & 14/7 41: South African female; 40s)
- 'Because it is logical' (9/6 31: South African female; 20s)

Of those respondents whose explanation did not agree with the statement ( $n=150$ ), 60% ( $n=90$ ) cited a religious rationale for their view. The most common explanation by these respondents was of God as creator, for example:

- 'God created us' (14/7 20: South African family group)
- 'God created everything' (16/6 12: South African female; 40s)
- 'We were all created by an all-powerful being. GOD' (1/5 47: South African male; teenager)

**Table 5:** Percentage of responses based on the conceptual framework using rationalising statements

Rationalising statement	Evolution acceptors	Evolution rejecters	Unsure
Naturalistic evidential	47		0
Naturalistic evolution	27		0
Naturalistic scientific	18		0
Creationist religious		60	0
Creationist evidential		33	0
Mixed rationale	1	3	7
Other	8	1	0

The next most common explanation by these respondents was a statement that they do not accept evolution (5%), for example:

- 'I am not convinced' (14/7 6: South African male; 70s)
- 'There is no proven link in evolution theory' (13/7 30: Mauritian male; 40s)
- 'We are special' (14/7 30: German male; 50s)

Along similar lines, some of the disagreeing respondents appealed to the Bible as the source of authority. For example:

- 'Bible says NO to evolution' (16/6 35: South African family group)
- 'I am a Christian and believe the Bible is the truth and God created us as human beings' (1/5 38: Asian couple)

A total of 33% of the respondents who disagreed with the statement ( $n=50$ ) cited some form of evidence or argument for creationism. Some examples of these explanations are:

- 'Don't believe in evolution. Believe that an ape is an ape, human being is a human being – cause all this things I have seen there, there is no full evidence to convince me. God is a creator.' (18/5 4: South African teacher accompanying school excursion)
- 'When God created man, he wasn't confused and started with a Ape. Why there aren't there and half ape half human species today?' (misconception; 1/5 39: South African male; 40s)

A mixed rationale was used by a small number of those who agreed or disagreed with the statement, and 7% of those who said they were unsure. Examples of such reasoning are:

- 'I think there must be evolution, but why don't we have half-ape people running around now? Have we no stopped evolving?' (misconception; 8/6 14: South African male; 20s)
- 'Yes – the genetic evidence and bodily structure proves the theory. No – people do accept the Adam & Eve theory' (16/6 23: South African male; 30s)

## Discussion

The findings presented above represent a self-selected sample of people interested enough to visit the Cradle of Humankind. While there are no clear indicators of numbers of museum visits by South Africans, there is a healthy domestic tourism industry with 25 million trips taken in both 2012 and 2013<sup>33</sup> which suggest that the sample surveyed can provide baseline data for future surveys.

The sample answers to Question 8 show how misleading the term 'Cradle of Humankind' is to visitors. While it is a useful branding tool for tourism purposes, it promotes a misconception about the origins of *Homo sapiens*. The current scientific consensus suggests modern humans originated in East Africa. It would appear that the proclamation of the Cradle of Humankind as a World Heritage site in 1999, and its subsequent promotion in South Africa, has strongly influenced the visiting South African public such that they consider it has the only claim to the name. From a scientific viewpoint, it is recommended that the Maropeng and Sterkfontein Visitor Centres need to situate the World Heritage Site within a broader African context, so that the visiting public understands two issues more clearly. Firstly, different species of early hominids were found throughout the African continent and may have overlapped in both space and time. Secondly, the earliest *Homo sapiens* fossil specimens were found in East Africa 160 000 to 120 000 years ago, rather than in the Cradle of Humankind in South Africa.

Questions 8 and 9 suggest that a small number of participants (2%) appear not to discriminate between the origin of humans and the origins of life. Such a notion would appear to be related to the concept of 'deep time' which is difficult for the layperson to comprehend. Both of the visitor centres attempt to explain geological time scales but a visitor's cursory glance at such an exhibit is unlikely to be internalised. Deep time

is a crucial concept in evolution, therefore, it is recommended that the visitor centres find ways to make visitors engage with the idea.

The findings from Question 10 show that 59% of the South African respondents accepted the concept of human evolution, while 37% disagreed. A South African online survey in 2014 ( $n=1002$ ) found an evolution acceptance rate of 49% for the statement 'Humans developed from previous species of animals'<sup>28</sup>). These surveys suggest that the South African public's acceptance of human evolution is lower than the 59% from visitors to the Cradle. There are a number of surveys on evolution conducted around the world, and it can be instructive to compare the findings from the current study with them. The 59% acceptance rate from the current survey is in line with a 2009 survey (sample size and selection criteria unknown) which found that 54% of the South African public accept 'that it is possible to believe in a God and still hold the view that ... human life evolved ... as a result of natural selection' (figures for Great Britain and Egypt were 54% and 45% respectively). However, only 6% South Africans accepted that 'life on earth, including human life, evolved over time ... God played no part' (Great Britain 38%; Egypt 2%), and 43% consider that life 'was created by a God and has always existed in its current form' (Great Britain 16%; Egypt 50%).<sup>34</sup> The Pew Research Centre<sup>29</sup> found in 2013 that 60% of American adults ( $n=1983$ ) agree that 'humans have evolved over time'. In the current study, the proportion of South African respondents who reject evolution is in line with the Pew percentage (37% vs 33%), while those who didn't answer is higher (16% vs 7%). Other studies have found that visitors to museums are more likely to endorse evolution as the explanation of human origins compared with the general public.<sup>14</sup> Given the self-selected sample of the current study, the findings do not represent the views of the South African population, but they provide baseline data for future possible comparison within a relatively educated segment of the population (as indicated by their occupation).

The rationalising statements of the participants who accepted human evolution (naturalistic evolution) suggest that the majority of them (74%) used their knowledge of scientific evidence and evolution to back up their conviction. While the extent of their knowledge was not probed, the findings suggest that these visitors were aware of some of the evidence for evolution. Other 'acceptors' (18%) appeared to rely on the status of science or logic as a rationale for their view (naturalistic scientific). One of the implications of this is that the visitor centres should aim to present clear evidence for human evolution, which might enable the visitors to clarify their own interpretations of the issue.

In contrast to the majority of evolution acceptors, only 33% of the non-acceptors cited evidence to support their disagreement with the statement (creationist evidential). The majority (60%) appealed to their own absolutes such as God as the creator, and the Bible as truth (creationist religious). This suggests different ways of thinking between the two groups, with those who agreed with human evolution citing evidence, while those who did not agree using religious-based statements. Gould<sup>35</sup> suggested the principle of NOMA, or 'non-overlapping magisteria' in which both science and religion each have legitimate domains of teaching authority. This might be a possible option that the visitor centres could adopt: presenting the science of human evolution, while also acknowledging that believers have religious views which need to be acknowledged. By presenting the views of faith groups alongside the scientific narrative, Maropeng and Sterkfontein could encourage visitors to think more deeply about the issues involved, and stimulate debate and discussion.

The responses from both the participants who accepted and those who rejected human evolution contain numerous misconceptions. Two examples will suffice here: humans evolved from apes or monkeys; humans are not evolving. These mirror more general misconceptions about evolution found in the literature<sup>36</sup> and indicate that if the visitor centres aim to provide more than a tourist experience, they need to actively identify the commonest misconceptions, address them, and ensure that guides are well-versed in changing visitors' misconceptions. This is particularly important in the light of the large number of school students who visit the Cradle of Humankind.



## Conclusions

The majority of the sample of visitors surveyed understood the term 'cradle' as an origin or birthplace of humankind. Although the instrument used was only able to access their knowledge at a fairly superficial level, the fact that over 84% of participants referred to humans, hominids or fossils suggests that they possessed some general knowledge of human ancestor fossil remains. However, 63% of the visitors surveyed appeared to be unaware that there are other areas of the world which potentially have better claims to the title 'Cradle of Humankind'. This unfamiliarity was particularly prevalent amongst South African visitors, and suggests that the visitor centres need to show that the area is one of several sites important for early human ancestors. Such provision of information would assist in making the public more aware of the tentative nature of scientific knowledge; that science itself is continually evolving rather than being a fixed body of knowledge. This is a goal of scientific literacy programmes internationally.<sup>37</sup>

The concept of deep time is one which most visitors are likely to be unfamiliar with, yet it is crucial for understanding the process of evolution. Additional visitor centre exhibits engaging visitors with the concept would likely assist visitors to comprehend this difficult idea.

Nearly 60% of the sample of visitors accepted that humans evolved from an apelike ancestor, while 25% did not accept this. There were statistically significant relationships between acceptance and nationality, and acceptance and population group, suggesting that South African visitors are less accepting of human evolution than their international counterparts. This has implications for the visitor centres as over a third of South Africans do not accept human evolution, yet this is the dominant narrative of the exhibits. In terms of the conceptual framework, the findings show that 'evolution acceptors' use different rationalising statements (i.e. evidence) from 'evolution rejecters' (i.e. religion). In order to promote deeper thinking about human origins, maybe the centres should juxtapose scientific explanations of human origins against religious explanations. The current scientific debates about the various possible human ancestors could also be presented, as well as examples of the compatibility between religious belief and acceptance of evolution.

Finally, one of the purposes of museums and visitor centres is to encourage visitors to think differently about things. Getting people to think through their own beliefs needs to be made more explicit if visitor centres in the Cradle of Humankind are to make a greater impact on the scientific literacy of the South African public.

## Acknowledgements

We thank the Palaeontological Scientific Trust and the National Research Foundation for funding the project, the staff of the Maropeng and Sterkfontein Visitor Centres for allowing access to the site for data collection, Jess Vujovic and Tamsin Hunt for data collection, data capture and assistance with coding, my colleagues Marissa Rollnick and Audrey Msimanga for commenting on draft versions of the paper, and the anonymous reviewers after submission to the journal.

## References

1. Zimmerman M. Why evolution is the organizing principle for biology. *Phi Kappa Phi Forum*. 2009;89(1):4–7.
2. Miller JD, Scott E, Okamoto S. Public acceptance of evolution. *Science*. 2006;313(5788):765–766. <http://dx.doi.org/10.1126/science.1126746>.
3. Dempster E, Hugo W. Introducing the concept of evolution into South African schools. *S Afr J Sci*. 2006;102(3/4):106–112.
4. Heddy BC, Sinatra GM. Transforming misconceptions: Using transformative experience to promote positive affect and conceptual change in students learning about biological evolution. *Sci Educ*. 2013;97(5):723–744. <http://dx.doi.org/10.1002/sce.21072>
5. Evans E. Teaching and learning about evolution. In: Diamond J, editor. *The virus and the whale: Exploring evolution in creatures small and large*. Arlington, VA: NSTA Press; 2005. p. 25–41.
6. Spiegel A, Evans EM, Gram W, Diamond J. Museum visitors' understanding of evolution. *Museum Soc Iss*. 2006;1(1):69–86. <http://dx.doi.org/10.1179/msi.2006.1.1.69>

7. Christie P. *The right to learn: The struggle for education in South Africa*. 2nd ed. Johannesburg: Ravan Press; 1991.
8. Lever J. Teaching and learning about evolution. In: James WG, Wilson L, editors. *The architect and the scaffold: Evolution and education in South Africa*. Cape Town: Human Sciences Research Council; 2002. p. 10–44.
9. Darwin C. *The descent of man, and selection in relation to sex*. London: Murray; 1871. <http://dx.doi.org/10.5962/bhl.title.2092>
10. Meredith M. *Born in Africa: The quest for the origins of human life*. London: Simon & Schuster; 2011.
11. Tattersall I. Human origins: Out of Africa. *P Natl Acad Sci USA*. 2009;106(38):16018–16021. <http://dx.doi.org/10.1073/pnas.0903207106>
12. Wolpoff MH, Hawks J, Caspari R. Multiregional, not multiple origins. *Am J Phys Anthropol*. 2000;112:129–136. [http://dx.doi.org/10.1002/\(SICI\)1096-8644\(200005\)112:1<129::AID-AJPA11>3.0.CO;2-K](http://dx.doi.org/10.1002/(SICI)1096-8644(200005)112:1<129::AID-AJPA11>3.0.CO;2-K)
13. Bell P, Lewenstein B, Shouse A, Feder M, editors. *Learning science in informal environments: People, places, and pursuits*. Washington, DC: National Academies Press; 2009.
14. Storksdieck M, Stein J. What they bring with them: Museum visitors' perspectives on evolution. *ASTC Dimensions*. 2006;(3/4):8–9.
15. Abraham-Silver L, Kisiel JF. Comparing visitors' conceptions of evolution: Examining understanding outside the United States. *Visitor Stud*. 2008;11(1):41–54. <http://dx.doi.org/10.1080/10645570801938434>
16. Spiegel A, Evans E, Frazier B, Hazel A, Tare M, Gram W, et al. Changing museum visitors' conceptions of evolution. *Evol Educ Outreach*. 2012;5(1):43–61. <http://dx.doi.org/10.1007/s12052-012-0399-9>
17. Scott M. The pleasures and pitfalls of teaching human evolution in the museum. *Evol Educ Outreach*. 2010;3(3):403–409. <http://dx.doi.org/10.1007/s12052-010-0252-y>
18. Scott M. Writing the history of humanity: The role of museums in defining origins and ancestors in a transnational world. *Curator Museum J*. 2005;48(1):74–89. <http://dx.doi.org/10.1111/j.2151-6952.2005.tb00155.x>
19. Scott M. *Rethinking evolution in the museum: Envisioning African origins*. Abingdon: Routledge; 2007.
20. Doering ZD, Pekarik AJ. Questioning the entrance narrative. *J Museum Educ*. 1996;21(3):20–23. <http://dx.doi.org/10.1080/10598650.1996.11510333>
21. Gundling T. Human origins studies: A historical perspective. *Evol Educ Outreach*. 2010;3(3):314–321. <http://dx.doi.org/10.1007/s12052-010-0248-7>
22. Evans S. SA ranks its maths and science second last in the world [article on the Internet]. c2013 [cited 2015 May 23]. Available from: <http://mg.co.za/article/2013-04-17-sas-maths-science-education-ranked-second-last-in-world>
23. Diamond J, Luke JJ, Uttal DH. *Practical evaluation guide: Tools for museums and other informal settings*. Lanham, MD: AltMira Press; 2009.
24. Falk JH, Heimlich JE, Vernon CL, Bronnenkant K. Critique of a critique: Do zoos and aquariums promote attitude change in visitors? *Soc Anim*. 2010;18(4):415–419. <http://dx.doi.org/10.1163/156853010X524361>
25. Falk JH, Storksdieck M. Using the contextual model of learning to understand visitor learning from a science centre exhibition. *Sci Educ*. 2005;89(5):744–778. <http://dx.doi.org/10.1002/sce.20078>
26. Ballouard J-M, Mullin SJ, Ajtic R, Brito JC, ElMouden EH, Erdogan M, et al. Factors influencing schoolchildren's responses to a questionnaire in wildlife conservation education. *Inter J Sci Educ*. 2015;37(3):469–483. <http://dx.doi.org/10.1080/09500693.2014.993000>
27. Dickinson D. *A different kind of AIDS: Folk and lay theories in South African townships*. Johannesburg: Fanele; 2014.
28. South African Press Agency. South Africans don't know much: Survey [article on the Internet]. c2014 [cited 2015 Feb 26]. Available from: <http://mybroadband.co.za/news/general/98316-south-africans-dont-know-much-survey.html>
29. Pew Research Center. Public's views on human evolution [article on the Internet]. c2013 [cited 2015 Apr 16]. Available from: <http://www.pewforum.org/2013/12/30/publics-views-on-human-evolution/>
30. Hatch JA. *Doing qualitative research in educational settings*. New York: Suny; 2002.

31. Krippendorff K. Content analysis: An introduction to its methodology. 3rd ed. Los Angeles, CA: Sage; 2013.
32. Cameron A. Cultural and religious barriers to learning in basic astronomy: A South African study [PhD thesis]. Johannesburg: University of the Witwatersrand; 2007.
33. Strategic Research Unit. Annual tourism performance report [report on the Internet]. c2013 [cited 2015 Mar 24]. Johannesburg: Strategic Research Unit; 2014. Available from: [http://www.southafrica.net/uploads/files/2013\\_Annual\\_Report\\_v8\\_01102014\\_\(1\).pdf](http://www.southafrica.net/uploads/files/2013_Annual_Report_v8_01102014_(1).pdf)
34. British Council. Results of British Council global education Darwin international survey. London: Ipsos Mori; 2009.
35. Gould SJ. Nonoverlapping magisteria. *Nat Hist*. 1997;106:16–22.
36. Gregory TR. Understanding natural selection: Essential concepts and common misconceptions. *Evol Educ Outreach*. 2009;2(2):156–175. <http://dx.doi.org/10.1007/s12052-009-0128-1>
37. Lederman N, Lederman J, Antink A. Nature of science and scientific inquiry as contexts for the learning of science and achievement of scientific literacy. *Int J sci educ*. 2013;1(3):138–147.



# Spoilage potential of a novel group of bacteria isolated from dairy products

## AUTHORS:

L. Ingrid Tsôeu<sup>1</sup>

Piet J. Jooste<sup>2</sup>

George Charimba<sup>1</sup>

Celia J. Hugo<sup>1</sup>

## AFFILIATIONS:

<sup>1</sup>Department of Microbial, Biochemical and Food Biotechnology, University of the Free State, Bloemfontein, South Africa

<sup>2</sup>Department of Biotechnology and Food Technology, Tshwane University of Technology, Pretoria, South Africa

## CORRESPONDENCE TO:

Celia Hugo

## EMAIL:

hugocj@ufs.ac.za

## POSTAL ADDRESS:

Department of Microbial, Biochemical and Food Biotechnology, University of the Free State, PO Box 339, Bloemfontein, 9301, South Africa

## DATES:

Received: 18 June 2015

Revised: 07 Aug. 2015

Accepted: 07 Aug. 2015

## KEYWORDS:

identification; spoilage; *Chryseobacterium*, *Empedobacter*; dairy

## HOW TO CITE:

Tsôeu LI, Jooste PJ, Charimba G, Hugo CJ. Spoilage potential of a novel group of bacteria isolated from dairy products. S Afr J Sci. 2016;112(1/2), Art. #2015-0227, 8 pages. <http://dx.doi.org/10.17159/sajs.2016/20150227>

© 2016. The Author(s).  
Published under a Creative Commons Attribution Licence.

Cold-tolerant bacteria, also known as psychrotrophic bacteria, are notorious contaminants of milk in the refrigerated dairy food chain. These organisms, especially the pseudomonads, may produce heat-resistant enzymes that are responsible for the breakdown of proteins and lipids in milk and dairy products. Such reactions result in a variety of defects in the raw or unprocessed milk that may affect the suitability of such milk for further processing. The enzymes produced may cause defects in long-life dairy products such as cheese, butter and long-life milk. In the present study, a range of 18 yellow pigmented psychrotrophic bacteria, collectively known as flavobacteria, were isolated from local dairy products. One aim of this study was to identify these bacteria to species level using molecular techniques. A second aim was to determine the spoilage potential of these organisms based on profiles generated by the BIOLOG system (that may relate to hydrolytic enzymes produced). Of the 18 isolates, 14 belonged to the genus *Chryseobacterium* while 4 were identified as *Empedobacter* isolates. The most active spoilage organisms in this group were shown to be *C. bovis*, *C. shigense* and *E. brevis*. These findings illustrate that enzymatically catalysed defects in dairy products should not be attributed solely to acknowledged psychrotrophic bacteria such as the pseudomonads, but that flavobacterial species may also be actively involved.

## Introduction

Extended cold storage times of raw milk at 2–6 °C have a significant influence on the natural microbial population. At these temperatures, the predominantly Gram-positive mesophilic aerobic bacteria are gradually replaced by Gram-negative and Gram-positive psychrotrophic bacteria.<sup>1,2</sup> Psychrotrophic bacteria are able to grow at 7 °C or less regardless of their optimal growth temperature.<sup>3</sup> In addition to this ability, psychrotrophic bacteria have the ability to produce heat stable extracellular and/or intracellular hydrolytic enzymes which may retain their activity even after the conventional heat treatment of milk.<sup>4</sup>

The genera *Pseudomonas*, *Aeromonas*, *Serratia*, *Acinetobacter*, *Alcaligenes*, *Achromobacter*, *Enterobacter* and *Flavobacterium*, with the predominance of *Pseudomonas* as the most frequent representative of Gram-negative psychrotrophic bacteria, have been isolated from raw milk.<sup>5</sup> Yellow pigmented psychrotrophs included in the bacterial family Flavobacteriaceae have the potential to decompose milk and dairy products.<sup>6</sup> These flavobacteria produce proteolytic enzymes that may survive pasteurisation<sup>7–9</sup> and cause bitter flavours, gelation of long-life heat treated milk<sup>10</sup> and the production of off-odours.<sup>11–13</sup> Jooste and Britz<sup>8</sup> found that the practical importance of dairy flavobacteria lies as much in their psychrotrophic growth and consequent proteinase production in refrigerated milk as in their contamination of milk via poorly sanitised pipelines and equipment.

While the taxonomy of the so-called flavobacterial group was in disarray for many years, this situation has stabilised<sup>14,15</sup> and more recently 114 genera have been included in the Flavobacteriaceae family.<sup>16</sup> Of the 114 genera in this family, 10 are associated with food, namely *Bergeyella*, *Chryseobacterium*, *Empedobacter*, *Flagellimonas*, *Flavobacterium*, *Myroides*, *Salegentibacter*, *Tenacibaculum*, *Vitellibacter* and *Weeksella*.<sup>17</sup> The latter reclassification and, in some cases erroneous classification of the flavobacteria in the past, makes interpretation of information about the incidence and role of flavobacteria in food deterioration a challenging task.<sup>18,19</sup>

Because of the latter changes in the flavobacterial taxonomy, it was decided to re-evaluate flavobacterial isolates from a previous study.<sup>14</sup> The aims of the study were therefore to more adequately identify flavobacterial strains isolated from dairy sources by 16S rRNA sequencing and phenotypic characterisation and then to estimate or predict the potential spoilage characteristics of the isolates by utilisation of the considerable range of substrates in the BIOLOG system.

## Materials and methods

### Revival of freeze-dried dairy isolates

The 18 isolates used in this study were isolated in a previous study, from raw cow's milk and butter samples from diverse regions of South Africa.<sup>14</sup> (Table 1). The isolates were preserved in the freeze-dried state and maintained at -20 °C. These freeze-dried cultures were revived by inoculation into nutrient broth (NB; Oxoid CM67, Thermoscientific, Basingstoke, United Kingdom) and incubated for 48 h at 25 °C. This was followed by streaking onto nutrient agar (NA; Oxoid CM003; Thermoscientific) and incubation for 48 h at 25 °C. Gram staining, oxidase and catalase tests<sup>20</sup> were performed on pure cultures to verify that the isolates were Gram-negative and oxidase and catalase positive, before further analysis.

### 16S rRNA sequencing

The DNA of the 18 isolates was extracted as previously described.<sup>21</sup> Polymerase chain reaction (PCR) amplification of the 16S rRNA gene was performed using an Eppendorf Mastercycler Temperature Gradient Personal Thermal cycler (Applied Biosystems, Johannesburg, South Africa) and the forward, 27F (5'-GAGTTTGATCCTGGCTCAG-3') and reverse, 1492R (5'-GGTACCTGTTCAGACTT-3') primers.<sup>22</sup> The PCR product was visualised on a 1% (w/v) agarose gel.

**Table 1:** Flavobacterial isolates used in this study

Isolate code <sup>14</sup>	Tentative grouping <sup>14</sup>	Source and region of isolation in South Africa
Hb8b	<i>Chryseobacterium balustinum</i> -like	Raw milk, Heidelberg
MK4	<i>Chryseobacterium balustinum</i> -like	Raw milk, Middelburg-Kriel
MK5	<i>Chryseobacterium balustinum</i> -like	Raw milk, Middelburg-Kriel
MK2b	<i>Chryseobacterium balustinum</i> -like	Raw milk, Middelburg-Kriel
PE17	<i>Chryseobacterium balustinum</i> -like	Raw milk, Port Elizabeth
Hbg2b	<i>Empedobacter</i> -like	Raw milk, Heidelberg route
Ss10a	<i>Empedobacter</i> -like	Raw milk, Steenkoolspruit
83.1	<i>Empedobacter</i> -like	Raw milk, Bloemfontein
D413	<i>Weeksella</i> -like	Butter, Bloemfontein
Dn13a	<i>Weeksella</i> -like	Raw milk, Balfour (Wits)
G222	<i>Weeksella</i> -like	Butter, Winburg
Hbg11a	<i>Weeksella</i> -like	Raw milk, Heidelberg route
NG7b	<i>Weeksella</i> -like	Raw milk, Nigel (Wits)
NG8a	<i>Weeksella</i> -like	Raw milk, Nigel (Wits)

The 1.5 kb bands obtained were excised from the agarose gel and purified using the DNA clean and concentrator kit according to the manufacturer's protocol (Zymo Research, Irvine, CA, USA). Sequencing was performed using both the forward and reverse primers with an ABI BigDye<sup>®</sup> Terminator v1.1 sequence cyclor (Applied Biosystems, Johannesburg, South Africa) according to the manufacturer's instructions. The newly determined 16S rRNA sequences (ca. 1500 bp) were compared to those available on the NCBI GenBank database<sup>23</sup> using the Basic Local Alignment Search Tool (BLAST). A phylogenetic tree was constructed using the neighbour-joining method in the MEGA 4 software<sup>24</sup> to determine the relationship of isolates in this study with those on the NCBI GenBank database.

### Phenotypic characterisation

In order to support the data obtained from the 16S rRNA sequencing, a range of phenotypic tests<sup>19,20,25</sup> were conducted using 24 h-old cultures on NA at 25 °C. The tests included pigment production, motility (flagellar type), growth in 0–5% (w/v) sodium chloride, growth at 5 °C, 25 °C, 37 °C and 42 °C; growth on  $\beta$ -hydroxybutyrate agar and MacConkey agar (Oxoid CM115, Thermoscientific); production of catalase, oxidase, phosphatase, DNase, urease, indole (Kovacs reagent, Merck 1.09293),  $\beta$ -galactosidase (ONPG) and reduction of nitrate; hydrolysis of esculin, gelatine (tube method) and starch; and acid production from carbohydrates in ammonium salts medium containing the following sugars: arabinose, cellobiose, D-fructose, D-glucose, lactose, maltose, mannitol, rhamnose, sorbitol, sucrose, trehalose, and D-xylose.

### Estimation of potential spoilage

For the estimation and prediction of potential spoilage characteristics, the isolates were streaked out and incubated at 25 °C for 24 h. The isolates were inoculated onto BIOLOG GN2 microplates (BIOLOG Inc., Hayward, California) according to the manufacturer's protocol. The 96-welled microplates contain 96 different oxidisable carbon sources which could be grouped into carbohydrates, amino acids, carboxylic acids and polymers and which were used for the estimation of the potential spoilage characteristics.

## Results and discussion

### Identification of isolates

The results of the 16S rRNA sequencing indicated that 14 of the 18 dairy isolates belonged to the *Chryseobacterium* genus with sequence

similarities between 94% and 99%, while four isolates belonged to the *Empedobacter* genus with sequence similarities between 96% and 98%. The phylogenetic relationships of the isolates to the type strains of *Chryseobacterium* species and *Empedobacter brevis* are shown in Figure 1.

Isolate Hb8b was closely associated with the type strain of *C. aquaticum* which was originally isolated from a water reservoir.<sup>26</sup> Isolate Hbg11a was closely associated with the type strain of *C. haifense* while five isolates (G222, MK5, D413, Dn13a and NG7b) were identified as *C. bovis*. The type strains of both *C. haifense* and *C. bovis* were isolated from raw milk in Israel.<sup>27,28</sup>

Two isolates (MK2b and 74.1) were identified as *C. indoltheticum*. The type strain of *C. indoltheticum* was originally isolated from marine mud.<sup>29</sup> Isolates MK4 and NG8a were identified as *C. shigense*. The type strain of *C. shigense* was originally isolated from a lactic acid beverage in Japan.<sup>30</sup> Isolate Dn2b was identified as *C. ureilyticum*. The type strain of this species was originally isolated from a beer bottling plant.<sup>31</sup> One isolate (MK8a) was identified as *C. indologenes*. Although *C. indologenes* is also known as a human pathogen, strains of this species have been isolated from food and the environment in previous studies.<sup>11,18,20</sup>

Although isolate Vv5a grouped closely with *C. ginsengisoli*, it could not be assigned to this species because of the low bootstrap value of 81 (Figure 1) when compared to the type strain. These data may suggest that isolate Vv5a could represent a new species. Isolates Ss10a, 83.1, Hbg2b and PE17 were identified as *Empedobacter brevis*.

Phenotypic analyses supported the 16S rRNA sequencing results by demonstrating typical phenotypic characteristics of the genera *Chryseobacterium* and *Empedobacter*.<sup>21</sup> All isolates were strictly aerobic, Gram-negative rods with yellow flexirubin or carotenoid pigments. All isolates were positive for the production of catalase, oxidase, phosphatase, DNase; growth at 25 °C and growth in 0 and 1% NaCl. All the isolates were negative for motility; growth in 5% NaCl and acid production from arabinose, cellobiose, rhamnose, sorbitol and xylose. The differential characteristics of the genera and species identified in this study are shown in Table 2.

Although the BIOLOG method can also be used for identification purposes, the only flavobacteria that formed part of the BIOLOG database were *C. gleum/indologenes*, *C. meningosepticum*, *C. indoltheticum*,



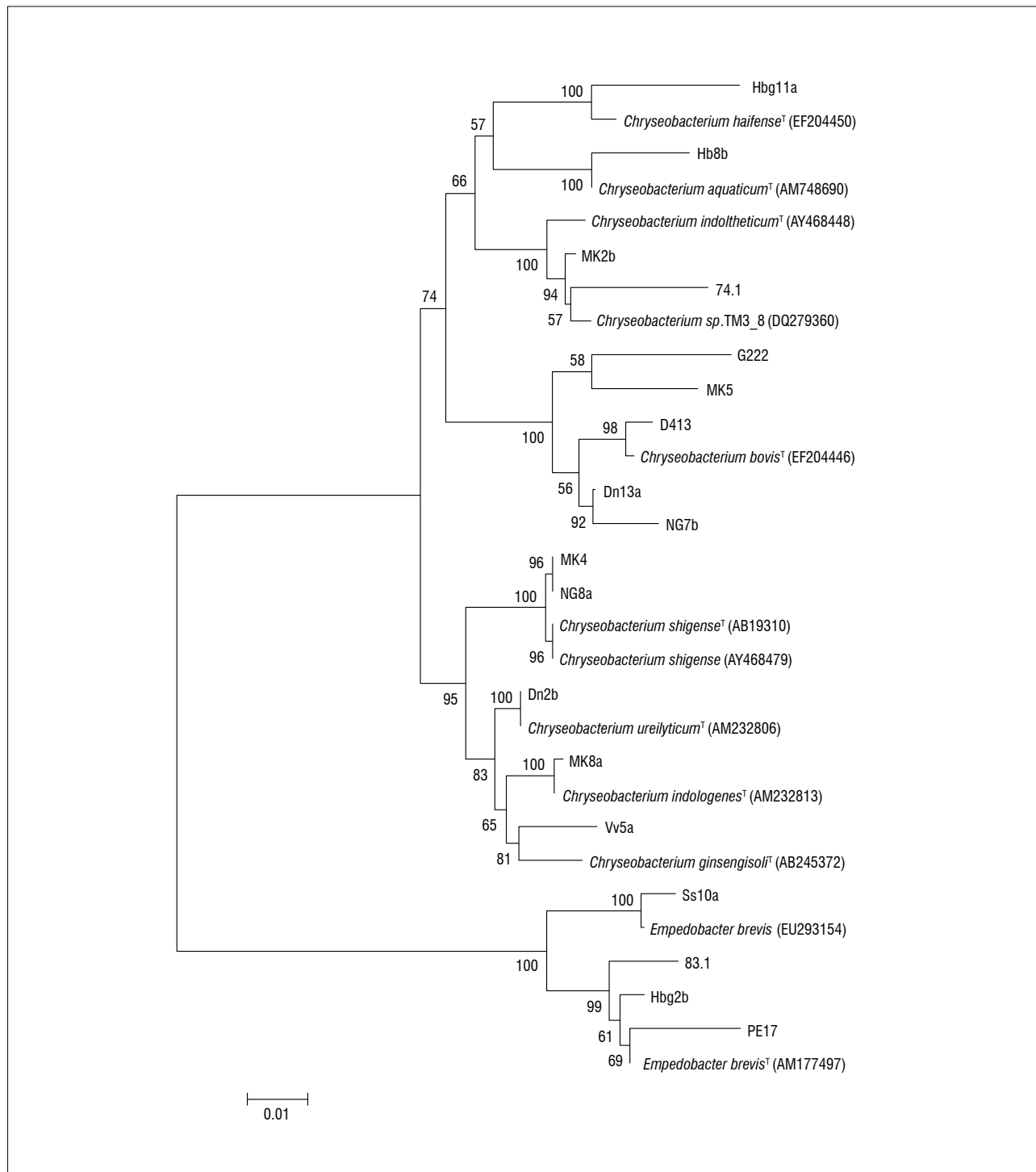
*C. scophthalmum*, *E. brevis* and CDC group II-E. The BILOG method could therefore only confirm that the isolates belonged to the Flavobacteriaceae family with grouping in either the *Chryseobacterium* or the *Empedobacter* genera (Table 3).

### Potential spoilage characteristics

#### Utilisation of carbohydrates

In this study, all the *Chryseobacterium* isolates (except *Chryseobacterium* strain Vv5a) and most of the *Empedobacter* isolates utilised glucose

(Table 4). The other carbohydrates were used in varying degrees by specific isolates. Bacterial types that are capable of oxidising a specific carbohydrate will use it as a carbon source to support growth. In the process, the carbohydrate is oxidised to CO<sub>2</sub> and water and, therefore, should not contribute to food spoilage. Only when the supply of simple, utilisable carbohydrates has been exhausted, will recognisable off-odours develop. These off-odours arise from the formation, and increase in concentration, of a variety of malodorous substances that include ammonia, amines, dimethylsulphide and hydrogen sulphide from amino acids.<sup>32,33</sup>



**Figure 1:** Neighbour joining phylogenetic tree showing relationships of the 18 dairy isolates from this study and closely related species from the GenBank database (accession numbers are given in parentheses). Bootstrap values are indicated at the nodes for 1000 replicates. The bar represents 10 substitutions per 100 nucleotides.

Interestingly, *Chryseobacterium* strain Vv5a did not utilise any of the carbohydrates tested in this study. This was in accordance with the 16S rRNA sequencing results, which showed that isolate Vv5a did not cluster with any of the type strains tested. These results lend impetus to placing the strain in a new species.

### Utilisation of amino acids

All the *Chryseobacterium* species (except strain Vv5a) and *E. brevis* isolates tested in this study were able to utilise glycyl-L-glutamic acid and/or L-glutamic acid (Table 4). The second most utilised amino acid was glycyl-L-aspartic acid. The oxidation/reduction of glutamic acid may yield acetic acid, butyric acid, carbon dioxide, ammonia and hydrogen.

Butyric acid and ammonia impart obnoxious (rancid) odours to food.<sup>34</sup> The incomplete metabolism of amino acids may result in putrescence, while the liberation of hydrogen sulphide from amino acids may result in a sulphide type of spoilage.<sup>35</sup>

On the more beneficial side, amino acids are the key precursors for ripened cheese flavour. They are metabolised by the action of amino acid-converting enzymes to aldehydes, alcohols, ketones, amines, acids, esters and sulfur-containing compounds, which all contribute to cheese flavour. While this action is desirable in some cheeses, it may be regarded as food spoilage in other products.<sup>36-41</sup> The off-flavours resulting from the above-mentioned compounds may range from fruity, malty, ripened-cheese to rancid, sweaty and putrid.<sup>42</sup>

**Table 2:** Differential phenotypic characteristics of the 18 dairy isolates in this study as identified using conventional methods

Characteristic	<i>Chryseobacterium aquaticum</i> (n=1)	<i>Chryseobacterium indoltheticum</i> (n=2)	<i>Chryseobacterium halitense</i> (n=1)	<i>Chryseobacterium bovis</i> (n=5)	<i>Chryseobacterium shigense</i> (n=2)	<i>Chryseobacterium ureilyticum</i> (n=1)	<i>Chryseobacterium indologenes</i> (n=1)	<i>Chryseobacterium</i> sp. (n=1)	<i>Empedobacter brevis</i> (n=4)
Production of:									
β-galactosidase	+ (-)	-	- (+)	1/5 (V)	-	-	-	-	1/4 (-)
Indole	+ (-)	- (+)	+	3/5 (-)	1/2 (+)	- (+)	- (+)	-	2/4 (+)
Urease	-	-	-	-	1/2 (-)	+	+ (-)	+	-
Nitrate reduction	-	-	+ (-)	1/5 (+)	-	-	-	-	2/4 (-)
Degradation of:									
Esculin	+	+	+	3/5 (+)	+	+	+	+	3/4 (-)
Gelatin	+	+	+	3/5 (ND)	+	+	+	+	2/4 (+)
Starch	+	-	-	1/5 (ND)	1/2 (+)	+ (-)	- (+)	+	3/4 (V)
Growth at:									
37 °C	- (+)	+ (-)	+	4/5 (+)	+ (-)	+ (-)	+ (-)	+	3/4 (+)
42 °C	-	-	-	1/5 (-)	1/2 (-)	-	-	-	-
Growth in NaCl:									
2%	+/-	+	+	4/5 (+)	+	+	+	+	3/4 (V)
3%	-	1/2 (+)	+ (-)	-	-	-	+ (-)	+	1/4 (V)
4%	-	-	-	-	-	-	+	-	-
Growth on:									
MacConkey agar	+ (-)	+	+ (-)	2/5 (-)	+ (-)	+ (-)	+	+	1/4 (+)
Acid from sugars:									
Fructose	- (+)	-	- (+)	-	- (+)	-	- (+)	-	- (ND)
Glucose	+	+	+	- (V)	+	-	+	+	+ (d)
Lactose	-	-	- (+)	-	-	-	-	-	- (ND)
Maltose	- (+)	- (+)	- (+)	1/5 (V)	-	-	- (+)	-	- (+)
Mannitol	-	-	-	-	-	+ (-)	-	+	- (ND)
Sucrose	+	-	- (+)	-	- (ND)	-	-	-	-
Trehalose	- (+)	-	-	-	-	-	- (+)	-	- (ND)

While the catabolism of amino acids is not a direct indication of hydrolytic breakdown of peptides or proteins, it is known that the flavobacteria are often psychrotrophic and proteolytic.<sup>7-9</sup> The majority of psychrotrophic bacteria have the ability to form hydrolytic, thermostable enzymes that break down the major constituents of milk, such as the protein.<sup>2</sup> The action of such enzymes destabilises the casein in milk through hydrolysis, and the result is the formation of a gel structure or coagulation of sterilised milk during storage.<sup>2</sup> The problem of the destabilisation of milk proteins leading to the precipitation or 'flocculation' of milk proteins during heat processing of milk has been identified as an important problem by the National Milkproducer's Research Workgroup of MilkSA.<sup>43</sup> Proteolytic enzymes of psychrotrophic bacteria have been cited as a possible contributing factor to this problem in certain milk producing regions of South Africa.

#### Utilisation of carboxylic acids

While acetic acid was metabolised by all the *Chryseobacterium* species (except *C. aquaticum*) and *E. brevis* in this study (Table 4), *C. bovis* and *E. brevis* were the only isolates capable of utilising  $\alpha$ -keto valeric acid. The metabolism of carboxylic acids by micro-organisms results in the coagulation of milk accompanied by sour and bitter tastes in milk.<sup>44</sup>

Formation of flavour compounds from branched chain  $\alpha$ -keto acids in fermented foods such as cheese is believed to be mainly an enzymatic process, because fermented product flavours are derived largely from amino acids.<sup>45</sup> Amino acids are mostly enzymatically transaminated in the bacterial cell to the corresponding  $\alpha$ -keto acid. These  $\alpha$ -keto acids are converted to various metabolites, such as the aldehydes and sulfur compounds which may result in off-flavours.<sup>46</sup>

**Table 3:** Identification of flavobacterial strains with the BIOLOG system after 24h of incubation and 16S rRNA similarity and identification

Isolate code	BIOLOG probability (%)	BIOLOG similarity	BIOLOG identification	16S rRNA similarity (%)	16S rRNA closest match
Hb8b	100	0.81	<i>Chryseobacterium gleum/indologenes</i>	96	<i>Chryseobacterium aquaticum</i>
MK4	83	0.57	<i>Chryseobacterium gleum/indologenes</i>	98	<i>Chryseobacterium shigense</i>
MK5	91	0.67	<i>Chryseobacterium meningosepticum</i> †	94	<i>Chryseobacterium bovis</i>
MK2b	–	–	–	98	<i>Chryseobacterium indoltheticum</i>
PE17	86	0.72	<i>Chryseobacterium indoltheticum</i>	97	<i>Empedobacter brevis</i>
Hbg2b	98	0.72	<i>Empedobacter brevis</i>	98	<i>Empedobacter brevis</i>
Ss10a	99	0.77	CDC group II-E subgroup	96	<i>Empedobacter brevis</i>
83.1	96	0.66	<i>Empedobacter brevis</i>	97	<i>Empedobacter brevis</i>
D413	88	0.51	<i>Chryseobacterium gleum/indologenes</i>	98	<i>Chryseobacterium bovis</i>
Dn13a	91	0.64	<i>Chryseobacterium scophthalmum</i>	98	<i>Chryseobacterium bovis</i>
G222	74	0.59	<i>Empedobacter brevis</i>	97	<i>Chryseobacterium bovis</i>
Hbg11a	84	0.65	CDC group II-E subgroup A	97	<i>Chryseobacterium haifense</i>
NG7b	-	-	-	97	<i>Chryseobacterium bovis</i>
NG8a	91	0.59	<i>Chryseobacterium gleum/indologenes</i>	99	<i>Chryseobacterium shigense</i>
74.1	82	0.56	<i>Chryseobacterium indoltheticum</i>	96	<i>Chryseobacterium</i> sp. TM3_8
Dn2b	85	0.78	<i>Empedobacter brevis</i>	99	<i>Chryseobacterium ureilyticum</i>
MK8a	99	0.70	<i>Chryseobacterium gleum/indologenes</i>	99	<i>Chryseobacterium indologenes</i>
Vv5a	77	0.64	<i>Chryseobacterium gleum/indologenes</i>	95	<i>Chryseobacterium ginsengisoli</i>

–, no results †*C. meningosepticum* = *Elizabethkingia meningoseptica*<sup>51</sup>

**Table 4:** Characteristics of the 18 dairy isolates indicating possible spoilage potential according to BIOLOG substrate utilisation

Characteristic	<i>Chryseobacterium aquaticum</i> (n=1)	<i>Chryseobacterium indoltheticum</i> (n=2)	<i>Chryseobacterium halifense</i> (n=1)	<i>Chryseobacterium bovis</i> (n=5)	<i>Chryseobacterium shigense</i> (n=2)	<i>Chryseobacterium ureilyticum</i> (n=1)	<i>Chryseobacterium indologenes</i> (n=1)	<i>Chryseobacterium sp.</i> (n=1)	<i>Empedobacter brevis</i> (n=4)
<b>Carbohydrates:</b>									
α-D-glucose	+	+	+	+	+	+	+	-	3/4
D-mannose	-	-	-	-	-	+	+	-	-
D-psicose	+	-	-	-	-	-	-	-	-
D-trehalose	+	-	-	1/5	1/2	-	-	-	-
Gentiobiose	+	1/2	-	2/5	+	+	+	-	-
Maltose	-	-	+	1/5	+	-	-	-	+
<b>Amino acids:</b>									
Glycyl-L-aspartic acid	-	1/2	+	2/5	1/2	-	-	+	+
Glycyl-L-glutamic acid	+	+	+	+	1/2	-	-	-	+
L-alanine	-	-	-	-	1/2	-	-	+	1/4
L-alanyl-glycine	-	-	-	1/5	1/2	-	-	-	2/4
L-asparagine	+	1/2	-	-	1/2	-	-	-	2/4
L-aspartic acid	-	-	-	1/5	1/2	-	+	-	2/4
L-glutamic acid	+	+	-	4/5	+	+	+	-	3/4
L-leucine	-	-	-	-	-	-	+	+	-
<b>Carboxylic acids:</b>									
α-keto butyric acid	-	-	-	-	-	-	-	-	2/4
α-keto valeric acid	-	-	-	3/5	-	-	-	-	3/4
Acetic acid	-	1/2	+	+	+	+	+	+	3/4
<b>Polymers:</b>									
α-cyclodextrin	+	1/2	-	3/5	1/2	-	-	-	3/4
Dextrin	+	+	-	3/5	1/2	-	-	-	3/4
Glycogen	+	-	-	1/5	-	-	-	-	-
Tween 40	-	-	-	1/5	1/2	-	+	-	2/4
Tween 80	-	-	-	1/5	1/2	-	+	-	2/4

-, negative result; +, positive result; 1/5, number of isolates positive for the specific reaction

#### Utilisation of polymers

α-Cyclodextrin and dextrin were the polymers most widely utilised by *C. aquaticum*, *C. indoltheticum*, *C. bovis* and *E. brevis* (Table 4). The polymers include polysaccharides. During food spoilage, complex di-, tri- or polysaccharides are usually hydrolysed to simple sugars before utilisation by microorganisms. Some of the metabolic products resulting from the oxidation of these polysaccharides include organic acids, alcohols and CO<sub>2</sub>, which may produce off-odours, sourness and bitter defects.<sup>34,47</sup>

Some strains of *C. bovis*, *C. shigense*, *C. indologenes* and *E. brevis* were able to utilise Tween 40 and Tween 80 (Table 4). Tween 40 and 80 hydrolysis serves as an indicator of lipase and other esterase activity<sup>48</sup> that can lead to spoilage of long-life dairy products during storage. Lipolysis by flavobacteria/chryseobacteria is known to break down fatty acids, resulting in rancid, butyric, bitter, soapy and astringent off-flavours and off-odours in some dairy products.<sup>49,50</sup>

When comparing the utilisation of the 22 substrates by the isolates tested in this study, *C. bovis*, *C. shigense* and *E. brevis* had the ability to



utilise 16 of the 22 substrates, which could indicate that these species have the potential to cause more pronounced types of spoilage than the other species tested. The frequency of utilisation of the 22 carbon sources by the remaining isolates were as follows: *C. aquaticum* (10/22), *C. indologenes* (9/22), *C. indoltheticum* (9/22), *C. haifense* (5/22), *C. ureilyticum* (5/22), *Chryseobacterium* sp. Vv5a (4/22).

## Conclusions

Flavobacteria previously isolated from dairy sources were shown in this study to belong to a range of *Chryseobacterium* species and to *Empedobacter brevis* by employing 16S rRNA sequencing and a variety of phenotypic tests. Although the BIOLOG system could not be used as an identification tool because of the limited database, it proved to be an excellent differentiation tool on the basis of the substrates utilised by the isolates. The BIOLOG system was also valuable in determining potential food spoilage characteristics of the isolates. The carbon sources most frequently utilised in the BIOLOG microplate system by the isolates were glucose, acetic acid, dextrin as well as amino acids such as glycyl-L-aspartic acid, glycyl-glutamic acid and L-glutamic acid. These may lead to spoilage characteristics such as off-odour production (e.g. fruity, malty, ripened-cheese, rancid, sweaty, putrid, sulphide-type), coagulation of milk, sourness and bitterness. Many of these spoilage characteristics result from the action of hydrolytic enzymes produced by the organisms. Proteolytic enzymes of psychrotrophic bacteria have been cited as a possible contributing factor to the important problem of milk 'flocculation' recently being encountered in certain milk producing regions of South Africa.

The flavobacterial species are believed to have the potential to cause spoilage defects in dairy products because they were able to utilise a wide range of compounds in the BIOLOG system. In this regard *C. bovis*, *C. shigense* and *E. brevis* were the most active and could utilise 16 of the 22 carbon sources even though varied reactions were shown where multiple strains of the organisms were tested. This was followed by *C. aquaticum* (10/22), *C. indoltheticum* and *C. indologenes* (9/22). Species that were least active included *C. haifense* and *Chryseobacterium* sp. Vv5a (4/22). These findings illustrate that enzymatically catalysed defects in dairy products should not be attributed solely to acknowledged psychrotrophic bacteria such as the pseudomonads, but that flavobacterial species may also be actively involved.

## Acknowledgements

This work was supported by funding from the South African National Research Foundation.

## Authors' contributions

C.H. was the project leader and responsible for experimental and project design; L.I.T. performed the experiments; P.J.J. made conceptual contributions; G.C. contributed to the interpretation and discussion of the biological results; L.I.T., P.J., G.C. and C.J.H. wrote the manuscript.

## References

- Lafarge V, Ogier J-C, Girard V, Maladen V, Leveau J-Y, Gruss A, Delacroix-Buchet A. Raw cow milk bacterial population shifts attributable to refrigeration. *Appl Environ Microbiol*. 2004;70:5644–5650. <http://dx.doi.org/10.1128/AEM.70.9.5644-5650.2004>
- Samaržija D, Zamberlin Š, Pogačić T. Psychrotrophic bacteria and milk and dairy products quality. *Mljekarstvo*. 2012;62:77–95.
- IDF Bulletin. Psychrotrophs in milk and milk products: E-doc 68. Brussels: International Dairy Federation; 1976.
- Chen L, Daniel RM, Coolbear T. Detection and impact of protease and lipase activities in milk and milk powder (review). *Int Dairy J*. 2003; 13:255–275. [http://dx.doi.org/10.1016/S0958-6946\(02\)00171-1](http://dx.doi.org/10.1016/S0958-6946(02)00171-1)
- Stepaniak, L. Psychrotrophic bacteria, bacteria other than *Pseudomonas* spp. In: Roginski H, Fuguay J, Fox PF, editors. *Encyclopedia of dairy sciences*, Vol. 4. London: Academic Press; 2002. p. 2345–2351. <http://dx.doi.org/10.1016/B0-12-227235-8/00407-7>
- Cousin MA. Presence and activity of psychrotrophic microorganisms in milk and dairy products: A review. *J Food Protect*. 1982;45:172–207.
- Gilmour A, Rowe MT. Microorganisms associated with milk. In: Robinson RK, editor. *Dairy microbiology*. London: Applied Science; 1981. p. 35–75.
- Jooste PJ, Britz TJ. The significance of flavobacteria as proteolytic psychrotrophs in milk. *Milchwissenschaft*. 1986;41:618–621.
- Venter H, Osthoff G, Litthauer D. Purification and characterization of a metalloprotease from *Chryseobacterium indologenes* 9a and determination of the amino acid specificity with electrospray mass spectrometry. *Prot Expr Purif*. 1999;15:282–295. <http://dx.doi.org/10.1006/prep.1998.1020>
- Meer RR, Baker J, Bodyfelt FW, Griffiths MW. Psychrotrophic *Bacillus* spp. in fluid milk products: A review. *J Food Protect*. 1991;54:969–979.
- Jooste PJ, Hugo CJ. The taxonomy, ecology and cultivation of bacterial genera belonging to the family Flavobacteriaceae (a review). *Int J Food Microbiol*. 1999;53:81–94. [http://dx.doi.org/10.1016/S0168-1605\(99\)00162-2](http://dx.doi.org/10.1016/S0168-1605(99)00162-2)
- Mielmann A, Hugo CJ, Jooste PJ. Food spoilage characteristics of *Chryseobacterium* species. Role and significance in food. Saarbrücken: VDM Verlag Dr. Müller; 2009.
- Bekker A. Growth and spoilage characteristics of *Chryseobacterium* species in milk [MSc dissertation]. Bloemfontein: University of the Free State; 2011.
- Jooste PJ. The taxonomy and significance of *Flavobacterium-Cytophaga* strains from dairy sources [PhD thesis]. Bloemfontein: University of the Orange Free State; 1985.
- Reichenbach H. Flavobacteriaceae fam. nov.: Validation of the publication of new names and new combinations previously effectively published outside the IJSB, List no. 41. *Int J Syst Evol Microbiol*. 1992;42:327–329. <http://dx.doi.org/10.1099/00207713-42-2-327>
- Parte AC. List of prokaryotic names with standing in nomenclature [document on the Internet]. c2015 [cited 2015 Aug 07]. Available from <http://www.bacterio.net>
- Hugo CJ, Jooste PJ. Culture media for food associated genera in the family Flavobacteriaceae. In: Corry JEL, Curtis GDW, Baird RM, editors. *Handbook of culture media for food and water microbiology*. Cambridge: RSC Publishing; 2012. p. 508–556.
- Vandamme P, Bernardet J-F, Segers P, Kersters K, Holmes B. New perspectives in the classification of the flavobacteria: description of *Chryseobacterium* gen. nov., *Bergeyella* gen. nov., and *Empedobacter* nom. rev. *Int J Syst Bacteriol*. 1994;44:827–831. <http://dx.doi.org/10.1099/00207713-44-4-827>
- Bernardet J-F, Hugo CJ, Bruun B. Genus X. *Chryseobacterium* Vandamme, Bernardet, Segers, Kersters and Holmes 1994, 829VP. In: Krieg NR, Staley JT, Brown DR, Hedlund BP, Paster BJ, Ward NL, et al., editors. *Bergey's manual of systematic bacteriology*. 2nd ed. New York: Springer; 2011. p. 180–196.
- Bernardet J-F, Nakagawa Y, Holmes B. Proposed minimal standards for describing new taxa of the family Flavobacteriaceae and emended description of the family. *Int J Syst Evol Microbiol*. 2002;52:1049–1070. <http://dx.doi.org/10.1099/00207713-52-3-1049>
- Labuschagne M, Albertyn J. Cloning of an epoxide hydrolase-encoding gene from *Rhodotorula mucilaginosa* and functional expression in *Yarrowia lipolytica*. *Yeast*. 2007;24:69–78. <http://dx.doi.org/10.1002/yea.1437>
- Lane DJ. 16S/23S rRNA sequencing. In: Stackebrandt E, Goodfellow M, editors. *Nucleic acid techniques in bacterial systematics*. Chichester: Wiley; 1991. p. 115–175.
- National Center for Biotechnology Information (NCBI). GenBank Overview. Rockville, MD: NCBI; 2011. Available from: <http://www.ncbi.nlm.nih.gov/genbank/>.
- Kumar S, Nei M, Dudley J, Tamura K. MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief Bioinform*. 2008;9:299–306. <http://dx.doi.org/10.1093/bib/bbn017>
- Hugo CJ, Segers P, Hoste B, Vancanneyt M, Kersters K. *Chryseobacterium joostei* sp. nov., isolated from the dairy environment. *Int J Syst Evol Microbiol*. 2003;53:771–777. <http://dx.doi.org/10.1099/ijs.0.02232-0>
- Kim KK, Lee KC, Oh H-M, Lee JS. *Chryseobacterium aquaticum* sp. nov., isolated from a water reservoir. *Int J Syst Evol Microbiol*. 2008;58:533–537. <http://dx.doi.org/10.1099/ijs.0.65491-0>

27. Hantsis-Zacharov E, Halpern M. *Chryseobacterium haifense* sp. nov., a psychrotolerant bacterium isolated from raw milk. *Int J Syst Evol Microbiol*. 2007;57:2344–2348. <http://dx.doi.org/10.1099/ijs.0.65115-0>
28. Hantsis-Zacharov E, Senderovich Y, Halpern M. *Chryseobacterium bovis* sp. nov., isolated from raw cow's milk. *Int J Syst Evol Microbiol*. 2008;58:1024–1028. <http://dx.doi.org/10.1099/ijs.0.65500-0>
29. Campbell LL, Williams OB. A study of chitin-decomposing micro-organisms of marine origin. *J Gen Microbiol*. 1951;5:894–905. <http://dx.doi.org/10.1099/00221287-5-5-894>
30. Shimomura K, Kaji S, Hiraishi A. *Chryseobacterium shigense* sp. nov., a yellow-pigmented, aerobic bacterium isolated from a lactic acid beverage. *Int J Syst Evol Microbiol*. 2005;55:1903–1906. <http://dx.doi.org/10.1099/ijs.0.63690-0>
31. Herzog P, Winkler I, Wolking D, Kämpfer P, Lipski A. *Chryseobacterium ureilyticum* sp. nov., *Chryseobacterium gambrini* sp. nov., *Chryseobacterium pallidum* sp. nov. and *Chryseobacterium molle* sp. nov., isolated from beer-bottling plants. *Int J Syst Evol Microbiol*. 2008;58:26–33. <http://dx.doi.org/10.1099/ijs.0.65362-0>
32. Dainty RH. Chemical/biochemical detection of spoilage. *Int J Food Microbiol*. 1996;33:19–33. [http://dx.doi.org/10.1016/0168-1605\(96\)01137-3](http://dx.doi.org/10.1016/0168-1605(96)01137-3)
33. Ellis DI, Goodacre R. Quantitative detection and identification methods for microbial spoilage. In: De Blackburn C, editor. *Food spoilage microorganisms*. Cambridge: Woodhead Publishing; 2006. p. 3–27. <http://dx.doi.org/10.1533/9781845691417.1.3>
34. Banwart GJ. *Basic food microbiology*. 2nd ed. New York: Van Nostrand Reinhold; 1989. [http://dx.doi.org/10.1007/978-1-4684-6453-5\\_3](http://dx.doi.org/10.1007/978-1-4684-6453-5_3)
35. Ayres JC, Mundt JO, Sandine WE. *Microbiology of foods*. San Francisco: WH Freeman and Company; 1980.
36. Aston JW, Creamer LK. Contribution of the components of the water-soluble fraction to the flavour of cheddar cheese. *N Z J Dairy Sci Technol*. 1986;21:229–248.
37. Crow VL. Metabolism of aspartate by *Propionibacterium freudenreichii* subsp. *shermanii*: Effect on lactate fermentation. *Appl Environ Microbiol*. 1986;52:359–365.
38. Crow VL. Properties of alanine dehydrogenase and aspartase from *Propionibacterium freudenreichii* subsp. *shermanii*. *Appl Environ Microbiol*. 1987;53:1885–1892.
39. Harper WJ, Wang JY. Amino acid catabolism in cheddar cheese slurries. I. Formation of selected products from alanine. *Milchwissenschaft*. 1980a;35:531–535.
40. Harper WJ, Wang JY. Amino acid catabolism in cheddar cheese slurries. II. Evaluation of transamination. *Milchwissenschaft*. 1980b;35:598–599.
41. Harper WJ, Wang JY. Amino acid catabolism in cheddar cheese slurries. III. Selected products from glutamic acid. *Milchwissenschaft*. 1981;36:70–72.
42. Engels WJM, Dekker R, De Jong C, Neeter R, Visser S. A comparative study of volatile compounds in the water-soluble fraction of various types of ripened cheese. *Int Dairy J*. 1997;7:255–263. [http://dx.doi.org/10.1016/S0958-6946\(97\)00003-4](http://dx.doi.org/10.1016/S0958-6946(97)00003-4)
43. Milk South Africa. Heat stability of reconstituted, protein-standardized skim milk powders [column on the Internet]. c2011 [cited 2015 Aug 07]. Available from: <http://www.milksa.co.za/research/research-column/heat-stability-reconstituted-protein-standardized-skim-milk-powders>
44. Urbach G. The flavour of milk and dairy products: II. Cheese: Contribution of volatile compounds. *Int J Dairy Technol*. 1997;50:79–89. <http://dx.doi.org/10.1111/j.1471-0307.1997.tb01743.x>
45. Smit BA, Engels WJM, Alewijn M, Lommerse GTCA, Kippersluis EAH, Wouters JTM, et al. Chemical conversion of  $\alpha$ -keto acids in relation to flavor formation in fermented foods. *J Agric Food Chem*. 2004;52:1263–1268. <http://dx.doi.org/10.1021/jf035147z>
46. Gao S, Oh DH, Broadbent JR, Johnson ME, Weimer BC, Steele JL. Aromatic amino acid catabolism by lactococci. *Le Lait*. 1997;77:371–381. <http://dx.doi.org/10.1051/lait:1997325>
47. Coulter TP. *Food: The Chemistry of its components*. London: Royal Society of Chemistry; 1984.
48. Harrigan WF, McCance ME. *Laboratory methods in food and dairy microbiology*. London: Academic Press; 1976.
49. Deeth HC, Fitz-Gerald CH. Lipolytic enzymes and hydrolytic rancidity in milk and milk products. In: Fox PF, editor. *Advanced dairy chemistry, Vol. 2: Lipids*. London: Chapman & Hall; 1995. p. 247–308.
50. Liao CH. *Pseudomonas* and related bacteria. In: De Blackburn C, editor. *Food spoilage microorganisms*. Cambridge: Woodhead Publishing; 2006. p. 507–540. <http://dx.doi.org/10.1533/9781845691417.5.507>
51. Kim KK, Kim MK, Lim JH, Park HY, Lee ST. Transfer of *Chryseobacterium meningosepticum* and *Chryseobacterium miricola* to *Elizabethkingia* gen. nov. as *Elizabethkingia meningoseptica* comb. nov. and *Elizabethkingia miricola* comb. nov. *Int J Syst Evol Microbiol*. 2005;55:1287–1293. <http://dx.doi.org/10.1099/ijs.0.63541-0>

